AD-770 550

NONVERBAL AND CULTURE FAIR PERFORMANCE PROCEDURES. II. INITIAL VALIDATION

Arthur I. Siegel, et al

Applied Psychological Services, Incorporated

Prepared for:

Office of Naval Research

September 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

UNCLASSIFIED
Security Classification

AD 770 550

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) APPLIED PSYCHOLOGICAL SERVICES, INC.	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE Nonverbal and Culture Fair Performance Procedures. II. Initial Validation	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Annual Report	
5. AUTHOR(S) (First name, middle initial, last name) Arthur I. Siegel, Brian A. Bergman, Joseph Lambert	
6. REPORT DATE September 1973	7a. TOTAL NO. OF PAGES 72 + vii 7b. NO. OF REFS 41
8a. CONTRACT OR GRANT NO. N 00014-71-C-0319 b. PROJECT NO. NR 153-334 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Personnel and Training Programs, Psychological Sciences Division, Office of Naval Research

13. ABSTRACT

The initial validation of a nonverbal, culture fair battery of tests for predicting performance of Navy machinist mates is described. The battery is based on the concept that ability to learn a miniature and representative aspect of a job can serve as a predictor of ability to learn the job as a journeyman. The battery was administered to 50 black and 49 white recruits who were below the minimal acceptable score for admission to the machinist mate school training, as measured by the usual Navy written tests. These recruits were placed on the job and their level of competance was measured through work sample performance test methods nine months later. It was possible to acquire certain criterion data for 29 of the black and 25 of the white subjects. The results indicated the performance battery to correlate higher with the performance criterion than the usual Navy tests. In a considerable number of cases, the "low aptitude" sample who had surpassed the minimal Navy test scores and who had entered the specialty after attending the Navy school for machinist mates.

1a

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Assessment						
Tests and Measurement						
Culture Fair Tests						
Personnel Prediction						
Differential Validity						
Performance Criteria						

ib

Nonverbal and Culture Fair Performance Prediction Procedures

II. Initial Validation

Arthur I. Siegel
Brian A. Bergman
Joseph Lambert

prepared for

Personnel and Training Programs
Psychological Sciences Division
Office of Naval Research

by

Applied Psychological Services, Inc.
Science Center
Wayne, Pennsylvania

under

Contract N00014-71-C-0319
NR 153-334

September 1973

IC

ABSTRACT

The initial validation of a nonverbal, culture-fair battery of tests for predicting performance of Navy machinist mates is described. The battery is based on the concept that ability to learn a miniature and representative aspect of a job can serve as a predictor of ability to learn the job as a journeyman. The battery was administered to 59 black and 49 white recruits who were below the minimal acceptable score for admission to the machinist mate school training, as measured by the usual Navy written tests. These recruits were placed on the job and their level of competence was measured through work sample performance test methods nine months later. It was possible to acquire criterion data for 29 of the black and 25 of the white subjects. The results indicated the performance battery to correlate higher with the performance criterion than the usual Navy tests. In a considerable number of cases, the "low aptitude" sample performed better on the criterion tests than persons in a control sample who had surpassed the minimal acceptable Navy test scores and who had entered the specialty after attending the Navy school for machinist mates.

ACKNOWLEDGMENTS

A number of other persons contributed substantially to the work here reported. Mr. Demetrius McDowell performed all the testing of the black subjects and interviewed black supervisors.

At the Navy Damage Control Training Center, CDR. P. May provided administrative and advisory support during the development of the criterion tests and during the work involved in deriving minimum acceptable and desirable scores on the criterion instruments.

The following persons provided necessary liaison with the various ships so that the necessary criterion performance data could be collected: CPO Willis (SERVLANT); LCDR Stapleford (PHIBLANT); CPO Stone (CRUDESLANT); LCDR Ryan (MINEWARFARE); CPO Jowers and Yoeman Smith (AIRLANT); CWO Hall, WO Poe, and CPO Beck (SUBLANT).

We acknowledge the substantial assistance provided by these persons and express our indebtedness for their gracious support.

Arthur I. Siegel
Brian A. Bergman
Joseph V. Lambert

APPLIED PSYCHOLOGICAL SERVICES, INC.
September 1973

TABLE OF CONTENTS

Table of Contents (cont.)

LIST OF TABLES

List of Tables (cont.)

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND PURPOSE

The present report is the second of a series in the area of "nonverbal and culture fair performance prediction procedures." The first report in the series (Siegel & Bergman, 1972) presented the logic of the present effort, the tests development, and initial findings. The present research program is based on a concept related to test "culture fairness" or "culture freeness," but is more properly associated with a "culture loaded" descriptor. By culture loaded, we mean performance prediction on the basis of instruments that have been loaded in the job culture of interest. The contention is made that a person's ability to learn a job sample, can be used to predict his ability to learn and perform the total job. Accordingly, the demonstrated ability to learn selected job aspects is employed as a predictor of ability to learn to perform the total job. The job sample tests (mini tests) involve no written materials and are relatively free from the usual confounding effects and influences of academic learning.

The specific purpose of the present effort is to investigate, in the Navy context, assessment methods, techniques, and procedures which are free from the biases ascribed to more conventional testing methods. This study does not focus on testing approaches which possess equal predictive validity for both high and low aptitude groups. Such a development is believed to be desirable, but not necessary. Alternatively, the assumption is made that normal Navy testing procedures (GCT + ARI + MECH) are sufficient for persons possessing high aptitude as measured by these methods. Persons who achieve high scores on these tests will not be affected to the extent that their progress in the military will be debilitated in any way. Alternatively, the Navy career of persons who score poorly on these usual Navy tests may be unduly affected by the cultural factors discussed in the previous and present report.

The specific research steps within the total investigation include:

1. Development of a sample of miniaturized job learning situations (tests) relevant to the machinist mate (MM) rating in the Navy. These miniaturized job learning situations are called training and evaluative situations in subsequent portions of this report.

2. Administration of these tests to a sample of low aptitude black and white recruits and assigning these recruits to predicted successful and predicted unsuccessful groups on the basis of their test scores.

3. Assignment of all persons sampled to Fleet jobs in the machinist mate rating.

4. Followup after the group has nine months of Fleet experience, to determine the degree of on-the-job success experienced.

5. Followup, after the group has 18 months of Fleet experience, to determine the degree of on-the-job success experienced.

To date, the first four of these steps have been completed. The remaining sections of Chapter I of this report include an extension of the literature review presented in the initial report of this series. This literature review contains discussions of the following relevant topics: (a) test and measurement error for various socioeconomic and racial groups, (b) personality, achievement, and self-concept factors affecting test scores, (c) effects of race of proctor/experimenter on test performance, (d) programs for disadvantaged and ghetto youth, and (e) cognitive style differences. The literature review provided in the prior report included, but was not limited to: (a) legal aspects of employment discrimination, (b) problems in establishing transethnic group test fairness, (c) studies into differential validity. (d) qualitative differences in intellectual functioning and performance, (e) performance comparative studies, and (f) motivational considerations.

Test and Measurement Error for Various Socioeconomic and Racial Groups

Garcia (1972) posited that intelligence tests should only be used as a measure of scholastic progress. In one study supporting this notion, Israeli Kibbutz children demonstrated that they could learn to do better on tests of intelligence. Garcia indicated that the idea of a single, general intelligence existing independently of environment is doubtful and possesses no social utility. According to Garcia, designers of intelligence tests incorporate assumptions which make the tests useless for comparing different ethnic and racial groups. Garcia concluded that intelligence tests are part of social conspiracy to promote the status quo.

Gael and Grant (1972) investigated the relevance of employment tests to actual job performance in an effort to generate culture fair tests that would predict potential for a service representative position. Minority ($n = 107$) and nonminority ($n = 193$) subjects, newly hired in the position, were given tests related to specially developed proficiency criteria. The results indicated that individual and composite test and criterion averages obtained by the two samples differed significantly, but that the validity coefficients were comparable. Regression equation comparisons indicated that common test standards could be used to evaluate minority and nonminority job applicants.

Green (1972) examined racial intelligence differences from the genetic and environmental points of view. Eleven hundred and twenty-seven Puerto-Rican subjects were placed into five groups along a skin color continuum (group 1 was white, group 5 was black, and the other groups were in between). Intelligence test results indicated a similarity between groups 1,2, and 3. Groups 4 and 5 scored lower than these groups. These results were said to parallel the social situation in Puerto Rico, where obvious blacks are victims of racial prejudice, while non-whites of lighter skin are accepted as white. Green compared the Puerto Rican situation to that of the United States, where the prejudice line is drawn between whites and non-whites (between groups 1 and 2). The conclusion drawn was that intelligence test scores follow the prejudice line more closely than the genetic line.

Mercer (1972) explored three problems: (1) what is mental retardation and how does the confusion of criteria affect the labeling of minority groups? (2) does the so called retarded I.Q. affect a person's ability to get along with others and to cope with the environment? (3) does the different cultural backgrounds of blacks and

Chicanos have anything to do with lower I.Q. scores? Tests and questionnaires were administered to retarded adults and children in a medium size city. It was concluded that: (1) persons of lower income groups are more likely to be labeled as retarded, (2) raw I.Q. scores are invalid without some test of environmental adaptation (many minority people classified as retarded are well adapted to their environment), and (3) I.Q. tests are really measures of the degree of indoctrination into the anglo culture.

The SPSSI (1969) statement on race and intelligence reaffirms the idea that racial differences in intelligence are not innate. Research has indicated that the more similar the background across white and black groups, the more similar the intelligence test scores. Because of various forms of subtle discrimination, blacks can not lead the same type of life as whites who possess comparable class backgrounds. Only when equality has existed for several generations can we begin to compare racial differences in intelligence test performance.

In another study (Williams, 1972), results of the Iowa Tests of Basic Skills were correlated with race and parental income level for public school children. It was determined that while race could account for some of the achievement decrement, extreme poverty (regardless of race) was a more significant variable in achievement decrement. In a similar study Southern and Plant (1971) found that young, preschool and kindergarten children of poverty stricken families demonstrated deficient general intellectual and language abilities.

Green and Rohwer (1971) examined the relationship between socioeconomic status (SES) and several academic and learning measures. They found that SES was related to a digit span test score, and to the Raven Progressive Matrices test score. SES was not related to paired associates learning. SES was also found to be related to long term school learning measures, an I.Q. measure, a complex problem solving measure, and an immediate memory task. No relationship was found between SES and a measure of short term learning.

Vineberg, Taylor, and Caylor (1970) and Vineberg and Taylor (1972a, 1972b) compared the job performance of men in five different specialties at several AFQT levels. They found that performance test scores were related to both AFQT scores and job experience. Job experience, though, accounted for considerably more of the performance test variance than did AFQT scores. Also, there was considerable overlap in performance test scores at different AFQT levels. These writers suggested that a large number of potentially good performers are lost to the Army because of low AFQT scores. In addition, the job performance test scores of whites and blacks were not significantly different even though blacks had lower average AFQT scores.

Ruch (1972) evaluated 20 tests in a business/industrial setting to determine whether or not the tests possess differential validity for blacks and whites. Separate statistics were computed for both blacks and whites. Analysis involved significance tests of homogeneity of regression lines, standard error, and intercepts. The results indicated no evidence of differential validity for regression line and standard error parameters. Significance data from the intercept parameter indicated

that black performance was overestimated, and that the tests were unfair to whites. Ruch concluded that following OFCC and EEOC guidelines involving common [illegible] scores and prediction procedures reduces employment opportunities for blacks.

Hunt (1971) reports that most studies find that the proportion of intelligence test score variance that is attributed to heredity is roughly 80 per cent. These studies have led to the belief that racial and ethnic differences in intelligence are fixed at birth. This notion is contrary to evidence derived from the plasticity of behavior research. In this research, behavior seems to be highly dependent upon manipulations of the environment. The plasticity studies indicate that early environment accounts for over 50 per cent of the behavioral variance. Hunt feels that heritability only tells the variance of genotypic intelligence for children reared in a specific static environment. It doesn't indicate the changes in intelligence that can take place among children reared in different environments with varying educational programs. When many of the socioeconomic and environmental differences are controlled across race and class membership, achievement and intelligence test score differences tend to disappear.

Humphreys (1972) discusses the various aspects of test fairness with respect to making judgments about individuals from differing groups. First, test decisions should be made only with maximally valid tests. Decisions should never be based on short tests. Second, Humphreys indicates that we can never achieve tests with perfect fairness due to regression problems. We can, though, fit the regression line to the data in cases in which the deviations from the regression line are small. Third, there is little difference in slopes and intercepts across various demographic groups such as indicated by most data. Use of a single regression line therefore causes no appreciable unfairness in drawing inferences about performance criteria. The amount of error, in fact, is generally less than the sampling errors of the regression coefficients. Finally, Humphreys contended that the same psychological principles apply to both whites and blacks since the differences across both groups have not been too large. If not, blacks and whites must be considered to represent different species.

Humphreys indicated that all of the aforementioned comments apply to achievement tests. Measures of intelligence, though, which depend upon the opportunity to learn, are highly unfair. Accordingly, tests are better at predicting behavior than at making inferences about theoretical constructs like learning ability.

Personality, Achievement, and Self-Concept Factors Affecting Test Scores

Katz (1969), in a critique of personality deficit, or cultural deprivation theories of underachievement, suggested that two factors account for most of the black achievement variance: (1) sense of environmental control," and (2) "school anxiety." Both have their source in early familial experiences,* and both seem to be modifiable by other (later) school experiences. According to Katz, other factors, such as the "father absent" and "n-achievement" hypothesis, tend to be unrelated to black school achievement. Katz also criticized the "cultural conflict" hypothesis, or the idea that competence in the black culture is irrelevant to the culture of the school. Research has shown, though, that the "culture conflict" hypothesis reduces to expectancy of attaining goals in a given culture, not in the choice of the culture to which one strives to achieve.

The three studies that are discussed in the following sections produced, essentially, incompatible results. All of these studies attempted to relate self-concept to either race or SES. One of the studies achieved positive results (Lefebre, 1971), one achieved inverse results (Soares & Soares, 1971), and one achieved negative results. Perhaps if investigators in this area would: (a) agree on a definition of self-concept, (b) use a common measure of self-concept, or (c) use a common research strategy, equivocal results could be avoided.

Lefebre (1971) administered the Tennessee Self-Concept Scale to 40 white and to 40 black junior high school students who were matched for age, I.Q., and SES. Blacks tended to score significantly lower on the following scales: (a) total positive self esteem ($p<.01$), (b) behavioral self-concept ($p<.01$), (c) personality integration ($p<.01$), (d) ethical self-concept ($p<.05$), (e) identity ($p<.05$), (f) self-satisfaction ($p<.05$), and (g) general maladjustment of self-concept ($p<.01$).

Soares and Soares, (1971) found that disadvantaged children view themselves and think that others view themselves more positively than do advantaged children. Secondary school students (both advantaged and disadvantaged) exhibited a diminished self-concept when compared with elementary school students. Soares and Soares suggested the possibility that this phenomenon is due to pressures of higher level education.

Getsinger, Kunce, Miller, and Weinberg (1972) attempted to relate three measures of self-esteem to SES, welfare status, race, educational advancement, and sex. A group of 198 sixth grade urban children was used as subjects. The results indicated that: (1) none of the measured personal-social characteristics were correlated with self-esteem as measured by the Coopersmith Inventory scores, (2) all measures but educational advancement ($r= .17$, $p<.05$) were unrelated to the Soares and Soares inventory, and (3) significant correlations were found between Ziller inventory scores and SES ($r= .18$), race ($r= .18$), and education ($r= .17$). Welfare status, SES, race, and education were all significantly interrelated, but sex was related only to education. These writers concluded that race and SES are inconsequently related to self-concept in the given age group.

*The effects of early experience on the academic achievement of blacks is more completely discussed in Siegel and Bergman (1972).

Abramson (1971) matched 22 ninth grade black subjects from an integrated school with 22 black subjects from a segregated school on academic achievement. Level of aspiration was measured by a digit-letter-substitution test. A Mann-Whitney U test indicated that the integrated students had significantly higher aspiration levels ($p<.05$) than the segregated students. A nonstatistically significant correlation of .18 was found between academic achievement and aspiration levels for the integrated group, while a statistically significant correlation of .52 ($p<.01$) was found between these same two variables for the segregated group. Abramson suggested that the integrated students are unrealistic goal setters, or perhaps their high aspiration is indicative of future academic achievement.

Veroff and Peele (1969) observed that black males who transferred from an all black school to a white school increased their achievement motivation ($p<.05$), while black males who remained in the all black school scored the same on achievement motivation. There were no differences with regard to black females across both types of schools.

Race of Proctor/Experimenter on Test Performance

Yando, Zigler, and Gates (1971) administered tests designed to assess social approach and avoidance, curiosity, and intelligence to 72 black and to 72 white "lower class" children. The examiners were: three "effective" white and three "effective" black teachers, and three "ineffective" white and three "ineffective" black teachers. The results of this experiment indicated that the performance of the children was influenced by individual variations in the personal characteristics of the adults' race. For example, achievement on intelligence tests was a function of the effectiveness of the teacher, and was unrelated to race. These influences were relatively constant for both black and white children.

Pucky and Banta (1972) examined the differences in the way in which black and white experimenters interact with black and white subjects, and the effect of the interaction on test achievement. Thirty-six Negro and 36 white preschool children were the subjects. Both white and black subjects obtained higher scores with white experimenters than with Negro experimenters. White experimenters were rated as providing a more positive social atmosphere than Negro experimenters for both groups of subjects.

Gould and Klein investigated the effects of black and white testers on the performance of black ($n= 46$) and white ($n= 38$) students in a racially mixed testing situation. These researchers found that black subjects performed as well with white test administrators as they did with black test administrators on both timed and untimed intellectual tests. On social, attitudinal, and personality measures, though, blacks presented themselves more positively when tested by a white than when tested by a black test administrator.

Watson (1972) studied the effects of several stressful interpersonal variables on the I.Q. scores of black children. The manipulated variables in this study were: (1) face-to-face relations with the test administrator, (2) knowledge of the test's purpose, and (3) knowledge that their scores would be compared with the scores of white children. Watson concluded that: (1) stress itself can lower I.Q. scores, (2) repressed agression in the presence of whites can lower I.Q. scores, and (3) white test administrators are particularly threatening to younger black children.

Programs for Disadvantaged and Ghetto Youth

Tobias, Busby, Greenfield, Goldberg, and Mead (1968) described the techniques used by several antipoverty agencies to prepare the hardcore unemployed for industry. Mobilization for Youth, Inc. views such factors as arriving to work on time as a job skill requiring training rather than as a problem for the employer. One method to condition a person to arrive at work on time is to allow a time range of arrival from 9:00 a.m. to 9:30 a.m., but to allow the early arrivals to use the new and better equipment and the late arrivals to use old equipment. Eventually 9:00 a.m. would be the expected arrival time. Mobilization for youth sees two problems in transferring from training to the job: (1) verbal difficulties, and (2) transferral of work done by mobilization for youth to the real job situation.

Youth in Action attempts to resocialize disadvantaged youth by training them for jobs, teaching them not to hate themselves, and because of this hate engage in destructive behavior. In this program, in order that their perspective will be increased, youths are instructed to unknown aspects of the culture such as theatre and museums.

The New York City Board of Education Job Counseling Centers posit that personalization of all interactions with the disadvataged is a must, because these individuals feel that they are "non-persons." At the job counseling centers: (a) the clinical atmosphere is lessened, (b) waiting is kept to a minimum, (c) doors are kept open, (d) dress is informal, (e) there is a group coffee hour every evening, and (f) counseling is available when needed.

Skill Advancement, Inc. deals with the underemployed who are already employed at the lowest level in industry. They train management as well as the employees in order to create a climate in which the worker can advance. These employees are given skill training, human relations training, communication training, and legal aid counseling.

The Kilmer Job Corps Center trains, clothes, houses, and feeds 1760 students at a cost of $5,200.00 per student. One feature of their program is a daily group dynamics session in which 14 boys get together and talk about their problems, or matters occurring during the day which they did not understand.

Clark (1968) indicated that industry must understand the following about ghetto youth: (1) he learns to survive early and on his own and what whites consider legitimate is not taken seriously by the disadvantaged youth, (2) he knows that he is a cast off from society and he does not expect anything to change, (3) he does not trust society's promises even when made by blacks, (4) he will try to take advantage of whatever gimmicks he learns, (5) he has insight into the immorality of society, (6) he is oversensitive to criticism, "because he's not sure that he is not actually inferior," (7) he will always test other's acceptance of him through hostility, withdrawal, and outlandish comment testing. Clark's recommendations to industry were: (1) offers made by industry to the disadvantaged must be serious, genuine, deliverable, (2) the disadvantaged youth must be trained in his real deficiencies: reading and math or he will see industry as insincere, thus confirming his expectations, (3) structural standards of performance, the same for white and black, are needed, (4) realistic rewards for achieving standards should be given frequently, (5) the black youth must have the same opportunity for upward mobility as whites, (6) the black youth wants to be accepted as an individual as are whites, and (7) the black want the same freedom of choice as whites.

Chappell, Brenner, Dammond, Harrison, Lee, Stone, and Reade (1968) discussed the programs of several companies to utilize the hard core unemployed. The Xerox Corporation's project "step up" involves a 19 week training program for disadvantaged ghetto residents. The time is equally divided between training and work. Participants are placed in regular jobs as soon as they pass the employment tests. Eighty-nine per cent of the first 16 enrolled in the program were able to complete it.

Lockheed's program (Chappell, et al, 1968) for the disadvantaged is based on the following tenets: (1) train for specific jobs, (2) develop self confidence by breaking jobs down and rewarding the learning of each step, (3) teach using demonstration, (4) require good work habits, (5) provide counseling, (6) recognize individual differences, (7) train to high standards, (8) household heads make the best trainees, and guarantee a job and pay the trainee during training.

In the Sandia Corporation (Chappell et al, 1968) sensitivity training and communications training was stressed with a sample of 23 Mexican-American women. After implementation of these techniques job satisfaction and performance were found to improve.

Westinghouse's (Chappell et al, 1968) recommendations for the training of the disadvantaged in industry are that there should be: (1) honesty in the program, (2) rewards during training, (3) individualized and relevant training, (4) dynamic training, (5) integration of counseling and training, (6) learning of the factory culture, (7) avoidance of paternalism; and (8) training of plant and peer personnel to accept and work with the disadvantaged.

Cognitive Style

Flaugher and Rock (1972) investigated the differing patterns of abilities among high school males of black, white, Mexican-American, and oriental ethnic groups on a multitest aptitude battery. The results indicated a significant similarity among the factor loading patterns across the different ethnic and racial groups. Flaugher and Rock concluded that essentially the same patterns of ability exist regardless of the ethnic identity of the examinee.

Fifth and sixth-grade students (n= 356) from a suburban school were tested on the Raven Progressive Matrices to examine the internal consistency and culturally related error patterns of the test. Differential group error patterns were found. This finding suggests that the lower performance levels of lower class subjects in general and the [illegible] performance of lower class blacks may have been due to perceptual rigid[illegible]er than a lack of reasoning ability (Bartlett, Newbrough, & Tulkin, 1972).

Wilcox (1971) found that 60 disadvantaged black college students exhibited a qualitatively different associative style from 42 white college students. The associations of the black subjects showed restricted word association hierarchies, a lack of unique associations, and a prevalence of opposite and syntagmatic associations in response to target words. Wilcox concluded that the word associations of blacks show a "developmental lag or deficiency" when compared with the word associations of whites. He suggested training in associative skills as one means of eliminating the differences found.

Hallahan (1970) discussed research evidence demonstrating that black disadvantaged children respond impulsively rather than attentively to environmental stimuli. He indicated blacks to possess a shorter attention span and to react with hyperactivity. This according to Hallahan, is due to restrictive child rearing practices and a disordered environment which prevents the subjects from learning to respond selectively and appropriately to stimuli. Thus, the behavior of the disadvantaged child was compared to that of the brain damaged child. Hallahan recommended training the disadvantaged child in attention skills and structuring the environment to increase concentration and decrease distraction from extraneous stimuli.

Summary

The research discussed in the present chapter supports the conclusion that a significant proportion of the variance of test score differences for different groups can be accounted for by environmental and training differences. The results reported herein do not warrant the conclusion that race of the experimenter has an effect upon student test performance. However the personal qualities of the experimenter/proctor appear to have some effect on test scores. The importance of training deprived youths in basic skills and work habits and of orienting industry to accept the underdeveloped were stressed in the section on programs for training underdeveloped youth. Finally, cognitive style has been criticized by some as an explanatory concept which accounts for the test score differences sometimes yielded by different groups. The use of differential training strategies is suggested as a means for eliminating test score differences among these different groups.

CHAPTER II

METHODS

The miniature aptitude test concept formed the basis for the present research effort. The miniature aptitude concept is embedded within a miniature training and evaluation situation. To develop a battery of miniature aptitude tests for the machinist mate rating, the tasks typically performed within the rating were dimensionalized using job analytic techniques. Short training sessions (15-30 minutes) were then built around the identified job dimensions. These training or learning sessions emphasized performance, observation, and practice rather than reading and writing. Once training was completed, the evaluation phase commenced. The evaluation, of course, consisted of the miniature aptitude tests which measured the amount learned in the training or learning phase. Each miniature test consumed no more than 5-20 minutes of administration time. Ninety-nine low aptitude Navy recruits participated in the training evaluations. The rationale for the procedure is that if the recruit can learn to perform the variety of job related tasks presented in the miniature training situations, he can also learn to perform similar tasks through on-the-job training in the Fleet. The details of these steps were described in the prior report (Siegel & Bergman, 1972) of this series. For purposes of continuity, these steps are also reviewed in subsequent sections of the present chapter.

The ninety-nine recruits were assigned to Fleet ships for work in the machinist mate rate. Nine months later, a set of job sample performance tests was administered to determine the predictive power of the miniature aptitude tests. The methods and results of this first validation constitute the principal focus of the present report.

The machinist mate specialty was selected as the vehicle for accomplishing the goals of the present study because this rate involves performance of tasks which are largely nonverbal in nature. In addition, this rate is one in which there is not a large number of blacks. Nonetheless, it is a rate which should be attractive to most recruits since it is adequately high on the informal prestige scale for various Navy jobs and because it offers the potential for learning skills which can lead to post Navy employment.

The normal entry into the machinist mate career field in the Navy is through the Navy A school for machinist mates. This school involves training in the fundamental skills and knowledges required for performance at entry level in the Fleet.

Subjects

The subjects were Navy recruits who were identified after initial Navy testing at the Great Lakes Naval Training Center. As a basic requirement for participation, a subject had to have "failed"* the entry tests for the Machinist Mate School. Ninety-nine recruits were so identified. Fifty of these recruits were white and 49 were black. Virtually all of the subjects were between 19 and 20 years of age. All had expressed a desire to enter the machinist mate rating during a career counseling interview. The subjects were not told that they were selected for a special study.

Miniature Job Sample Test Development

The Machinist Mate section of the Navy Manual of Qualification for Advancement in Rating (NAVPERS 18068) was consulted as a first step in the construction of the miniature training and evaluation situations. Those practical behaviors required for advancement to level E-4 were extracted. Several of these behaviors were combined because of their similarity. The next step involved a meeting with five Master Chief Machinist Mates and one Warrant Officer at the Great Lakes Naval Training Center. During this meeting, a final list of behaviors, which were held to be adequately representative of the most frequently performed or critical tasks of the journeyman level machinist mate, were agreed on. The six behaviors identified were:

1. ability to identify and use hand tools common to the job
2. ability to perform maintenance and to read meters and guages accurately when under some degree of distraction, or when attention sharing is involved
3. ability to make simple repairs in pressure lines
4. ability to perform simple troubleshooting and systems analysis in pressure systems
5. ability to operate equipment common to rate
6. ability to assemble and disassemble common high failure frequency items

These tasks formed the basis for the miniature training and evaluation situations. Several Master Chief Machinist Mates then served as technical consultants during the actual lesson preparation phase of the project. This procedure resulted in the construction of six tests and the associated training situations. These training and evaluation situations reflected samples of the six most critical and/or frequently performed behaviors, as listed above, or the entry level machinist mate.

*A recruit must exceed a combined General Classification Test (GCT), Arithmetic (ARI), and Mechanical (MECH) score of 156 to be eligible for entry to the machinist mate "A" school.

Each situation contained two parts. The first part, a training phase, usually involved a "show and tell" learning situation of 15 to 30 minutes. Reading and writing ability were not required in any of these lessons although some of the tests involved the ability to tell time and the ability to read numbers. These are considered to be preliterate requirements.

Equipment Use and Nomenclature

In the equipment use and nomenclature training and evaluation situation, the goal was to determine whether or not the recruit could learn the names and uses of all of the equipment and material involved in breaking-making and flange. The assumption was made that if a sailor could learn the names and uses of the tools and materials involved in this situation, he would also be able to learn the names and uses of other equipment used on the job. After a tape recorded introduction, the instructor demonstrated how to break and make a flange. During this demonstration, the use and name of each piece of equipment was discussed by the instructor. For example, the instructor would hold up the tool and say "This is a _____." When the tool was returned to the instructor he would demonstrate its use in the flange repair situation. When the demonstration was completed, a 25 item true-false test was orally administered to the recruits. For each item in this test, the instructor held up an object and ascribed a name or use to it. The recruits then indicated whether the name or use given by the instructor was true or false by encircling either the word "true" or the word "false" next to the item number on their answer sheets.

Gasket Cutting and Meter Reading

The gasket cutting and meter reading training and evaluation situation was designed to investigate ability to learn a maintenance task and to perform when some degree of attention sharing is involved. This situation was also designed to sample the vigilance situation in which the machinist mate, on the job, must monitor the state of various equipment systems while he performs other tasks. After a tape recorded introduction, the subjects were taught, through demonstration, how to make a gasket using a flange, a ball pean hammer, asbestos gasket material, and some bolts. On completion of the lesson, the subjects were given a 10 minute gasket making practice session. During the practice sessions, the instructor circulated among the recruits and gave assistance when required.

Next, the subjects were taught how to read a pressure meter and: (1) how to log the time at which the pressure deviated from normal and, (2) how to know whether the pressure should be adjusted up or down (relative to a given nominal valve). A large clock with a sweep hand was placed in front of the room. The recruits were able to read the time from this clock. The tests for both gasket cutting and meter reading were administered together. That is, for a ten minute period, the subject had to observe and record from a meter while he constructed a gasket. The meter which each individual read was placed at his individual work station. Each meter was individually

driven so that there was no possibility for a subject, who noticed an abnormal condition at his station, to cue a recruit at another station of an abnormal condition at the second station. The signal presentations to all stations were equated for number, direction, and magnitude of deviation. Completion of these two tasks simultaneously introduced the attention sharing component into the measurement situation.

The gasket making was scored through a checklist which was completed by obervers as the recruits performed the task. The scoring checklist included items on adherence to correct methods, care and use of tools, adherence to safety precautions, and adequacy of the final gasket. The meter reading test was scored the basis of: (1) number of abnormal conditions correctly noted, (2) correct indication of the direction of the necessary pressure adjustment (up or down) required to restore the system to normal, and (3) precision of the log entry for time of deviation from normal.

Trouble Shooting

The objective of the trouble shooting miniature training and evaluation situation was to test the recruit's ability to learn to perform simple systems analysis and trouble shooting on a hydraulic system. A simulated pressure system was used as the apparatus for both the training and testing aspects. After a taped introduction, the recruits were taught how the pressure system operates. The apparatus consisted of a series of color coded interchangeable gears which were so interconnected that a simulated pumping system was driven. A set of valves controlled the flow within the simulated pumping system. Accordingly, to diagnose a malfunction in the system, the recruit was required to comprehend such elementary relationships as: (1) the effects of gear size on pump speed/rate of flow, (2) the effects of direction of gear rotation on flow, (3) how differential gearing can produce changes in output rate, and (4) the effects of valve and pump function on system operation.

A series of light indicators was used to signal adequacy of flow at various parts of the simulated system. The recruits' task was to observe the light indicators, to determine whether any malfunction existed, and to note the cause of the condition.

In the training situation, various malfunction situations were presented, and the recruits were taught what parts of the system needed adjustment in order to correct the problem and the logic for the correction. After training, the subjects were presented with a series of practice trouble shooting situations. For each problem, both the correct answer and the reason for it being correct were discussed.

Twelve problems were orally presented to the students in the test situation. Each problem presented a given malfunction indication. The subjects were required to identify, by encircling a number, the system component which would cause the given malfunction indication.

Equipment Operation

In the equipment operation miniature training situation, the recruit was taught to start up and shut down a motor and pump apparatus. The students were required to learn a 33 step procedure, including several safety precautions. Each subject was then given the opportunity to practice starting up and shutting down the apparatus. After practice, a checklist type performance test was administered. Scoring was accomplished during the subject's performance and was based on adherence to correct procedures and observance of safety precautions.

Assembly

In the assembly training and evaluation situation, the recruits were taught and tested on the assembly of a gate valve from its component parts. First, a demonstration of the correct assembly procedure was presented. This demonstration was followed by a practice session in which the students were allowed to assemble the valve themselves. The instructors observed the students during this practice period and helped them, as required. After the practice session, each recruit was individually tested on his ability to assemble the valve. Again, scoring was through the checklist procedure.

Pass Fail

A subject was considered to have "passed" the miniature training and evaluation battery if he scored "average" or better on the trouble shooting test and "average" or better on two of the five remaining tests. These subjects were assigned to a "predicted successful" category and randomly assigned to ships in the Fleet for work in the machinist mate rate. Subjects not meeting the "pass" criterion were assigned to a "predicted fail" category. These subjects were also randomly assigned to ships in the Fleet for work in the machinist mate specialty. The results indicated that there were no differences in the proportion of whites and blacks passing these tests.

The rationale behind the choice of passing scores was that the trouble shooting miniature training and evaluation situation was cognitive in nature, while the remaining situations involved the learning of manipulative procedures. Hence, those in the "predicted successful" group demonstrated some cognitive as well as some manipulative/procedural skill learning ability in the miniature job learning situation.

The score of each recruit in the sample on the GCT, ARI, MECH, and CLER tests of the Navy classification battery was provided by recruit classification personnel at the Great Lakes Naval Training Center. These data, presented in our earlier report (Siegel and Bergman, 1972), indicated fairly close agreement between the white and black groups on these tests.

Instructors and Test Administrators

Two instructors/test administrators managed each training and evaluation session. One instructor/test administrator was a retired, black, Chief Petty Officer. The second instructor/test administrator was a white psychologist from the Applied Psychological Services' staff. The assigned duties of the black Chief Petty Officer were:

1. taping the introduction to each session
2. conducting the training segment of each lesson
3. administering the tests to black recruits

This instructor was thoroughly trained in the content he was to present, the teaching methods he was to employ, and the test procedures prior to implementation of the present program.

The duties of the white psychologist were:

1. organization and direction of the mini job sample learning and testing program
2. training the black instructor/test administrator
3. assisting the black instructor/test administrator in his lesson presentations
4. testing the white recruits

A black instructor/test administrator was used because his education, personality, verbal inflection, and method of treating the subjects were of such a nature that he could easily be identified with and understood by the black recruits.

Setting

All of the miniature training and evaluation sessions were conducted in a large classroom provided by the Machinist Mate School, Great Lakes Naval Training Center. This classroom contained 12 student desks and six work tables of various sizes. The lighting, ventilation, temperature, space, and privacy were considered optimal for this type of study.

All sessions began at 0720 hours and were completed by 1500 hours. Between 6 to 12 recruits were involved during each training and evaluation day. The total training, practice, and testing was completed for any one group within the time indicated.

Questionnaire

A 36 item personal background questionnaire was constructed by Applied Psychological Services to measure various facets of cultural differentiation including: (a) need achievement, (b) home environment, (c) school environment, and (d) other demographic variables. This questionnaire was administered to all recruits in the miniature training and evaluation sample. It was considered that these cultural factors could moderate learning ability to the extent that the correlations between the miniature evaluation test scores and the ultimate performance criteria would be lowered.

In addition, Applied Psychological Services administered the same questionnaire to a control group of machinist mate A school recruits, who had met or surpassed the required GCT, ARI, and MECH test score for the machinist mate rating.

The questionnaire administered to the miniature training and evaluation sample was factor analyzed. Nine factors, accounting for 46 per cent of the variance, were extracted. These factors were called: self-esteem, environmental stimulation, reading habits, educational attainment, educational initiative, parental interest, monetary deprivation, educational encouragement, and urbanity. Our earlier report in this series (Siegel & Bergman, 1972) presents the actual items and factor loadings for each factor.

In order to test whether or not the low aptitude group differed significantly on the cultural differentiation factors from the A school group, factor score means for the high aptitude (A school) group and for the low aptitude group were calculated. Tests ("t" tests) were conducted between the group mean scores for each factor. For eight of the nine cultural factors, the low aptitude group differed significantly from the high aptitude group.

Finally, in order to determine if the miniature training and evaluation tests were less contaminated by cultural factors than the Navy qualification tests, correlation coefficients were calculated between the Navy qualification test scores, the miniature job sample tests, and the questionnaire factor scores for the 99 recruits in the "low aptitude" sample. Only one of the correlations between the miniature training and evaluation tests and the cultural factors was statistically significant. Alternately, seven of the correlations between the Navy qualification tests and the cultural factor scores were statistically significant. These data support a contention that the mini tests are less culturally loaded than the usual Navy qualification tests.

Interview

On completion of the miniature training and evaluation situations, each recruit was interviewed by one of the instructors/test administrators in order to obtain reactions to the entire training and testing program. The subjects were asked to compare the tests and training they received in the present program with other types of tests and training they received in the past. All subjects were encouraged to respond freely and openly to the interviewer. The interview questions were constructed in a manner permitting quantitative and qualitative analysis of the interviewee responses.

The results of this interview analysis are summarized in the following paragraphs*

Seventy-eight per cent of the recruits thought the training portion of the program was better than the training they had received in other programs. Twenty per cent of the subjects thought the training portion of the program was the same as the training they had received in other training programs. Only one per cent of the subjects thought that the training they received was worse than that of other training programs. These results tend to support the emphasis placed on performance and manipulative skills during the learning sessions rather than on reading and writing.

Eighty-six per cent of the subjects thought the miniature job learning tests were "better than" paper and pencil tests. Twelve per cent of the subjects thought the program tests were the "same as" paper and pencil tests. Again, only one per cent thought the program tests were "worse than" paper and pencil tests. These results support the use of performance oriented tests which require little or no reading.

Finally, 98 per cent of the subjects indicated that they enjoyed participating in the training and evaluation program, while only two per cent indicated that they did not enjoy it.

These responses and the results of the first two questions permit the conclusion that the low aptitude Naval recruits show an overwhelming preference for the training and evaluation program, as employed here, over the more traditional testing approaches.

---

* A more complete interview analysis was presented in the previous report of this series (Siegel & Bergman, 1972).

CHAPTER III

CRITERION DEVELOPMENT

Quite obviously, freedom from cultural bias, equivalence for both black and white groups, and the like represent necessary but not sufficient ingredients for an assessment approach. The predictive validity of the miniature job learning tests must also be demonstrated. To this end, the recruits in the "low aptitude" sample were followed up after they had served 6 to 9 months in the machinist mate specialty. The followup attempted to measure, through the work sample performance test approach, the ability of the men to perform various aspects of the job of the machinist mate. These criterion data were supplemented with supervisor ratings and supervisory interview data. In this report, we place primary emphasis on criterion referenced followup tests and only minimum emphasis on the supervisory ratings and interview data. It is known that supervisors often rate black job incumbents lower than whites. Flaugher, Campbell, and Pike (1969) found that white supervisors rated a group of Negro incumbents one half a standard deviation lower than Negro supervisors. Negro supervisors, though, did not rate white incumbents higher or lower than white supervisors.

Nevertheless, supervisory evaluative data are consi to be of interest in the present context.

Criterion Tests

To develop the criterion tests for use in this study. Applied Psychological Services' personnel first reviewed portions of NAVPERS 18068A relevant to the machinist mate rate. Several proposed ideas for criterion referenced six month Fleet performance tests were extracted. The next step was to elaborate on these testing possibilities with experts in the machinist mate rate. The experts in this situation were primarily needed to supply scorable testing items, fruitful testing suggestions, and statements of criteria for "acceptable" work on each criterion objective after six months experience as a machinist mate striker. Four such experts were made available by the Naval Damage Control Training Center, Philadelphia.

These experts agreed that the following practical performance items would constitute an adequate test of the ability of a machinist mate striker, with six months Fleet experience, to meet normal performance expectations:

1. standing messenger watch
2. making and breaking a flange
3. packing a valve
4. procedures in common malfunction and in emergency situations
5. knowledge of use and names of common equipment and tools
6. general alertness and common sense in the work situation

Three separate meetings were held to isolate, derive, and define the performance objectives and the methods for measuring performance on these objectives. As a result, six nonverbal performance tests were defined. Each of these is described categorically below.

Messenger Watch

The messenger watch examination tested examinee's ability to record accurately data and to determine malfunctions indicated by the data. Pictorial simulations of throttle board situations were presented, one at a time, to the examinee. The examinee was required to record accurately data from the meters and gauges depicted on the throttle board simulation. The examinee was also required to report to the examiner those throttle board readings which indicate an abnormality or malfunction in the system.

Thus, each examinee received two scores. One score was based on his degree of accuracy in recording data from the simulated throttle board pictures; the second score indicated his ability to detect malfunctions on the basis of given data.

The directions to the messenger watch test, which were given orally to each examinee prior to testing, are:

> "THIS IS A TEST OF YOUR ABILITY TO COPY NUMBERS FROM METERS AND GAUGES ONTO A LOG SHEET. THIS IS ALSO A TEST OF YOUR ABILITY TO TELL US IF THERE IS SOMETHING WRONG WITH THE GAUGE OR METER READING. YOU WILL BE PRESENTED WITH A SERIES OF PICTURES SHOWING COMMON METERS AND GAUGES FOUND ON A THROTTLE BOARD. WE WANT YOU TO RECORD THE READING ON EACH OF THESE METERS AND GAUGES ON A LOG SHEET WHICH WE WILL GIVE YOU. MAKE YOUR READINGS AS ACCURATELY AS POSSIBLE. ALSO, WE WANT YOU TO TELL US WHICH METERS AND GAUGES INDICATE AN OUT OF NORMAL CONDITION WHICH SHOULD BE REPORTED. THE PICTURES YOU WILL SEE SHOW A 600 POUND STEAM SYSTEM. THIS MEANS THAT THE MAIN STEAM, STEAM TO TURBINES, AND AUXILIARY STEAM SHOULD ALL READ ABOUT 600 POUNDS. ALL OTHER METERS AND GAUGES ARE READ THE SAME AS ABOARD YOUR SHIP. WE ONLY WANT YOU TO LOG YOUR READINGS FOR THE EXAMPLE AND FOR THE FIRST FOUR TEST PICTURES. FOR THE REMAINING PICTURES WE JUST WANT YOU TO REPORT ANY PROBLEM THAT YOU FIND WITH ANY OF THE READINGS. YOU WILL HAVE TWO MINUTES TO LOG THE READINGS FOR EACH OF THE FOUR TEST PICTURES, AND YOU WILL HAVE 30 SECONDS TO TELL US WHAT'S WRONG WITH EACH PICTURE. NOW, LET'S DO THE EXAMPLE TOGETHER."

A sample item from the messenger watch test is presented as Figure 3-1.

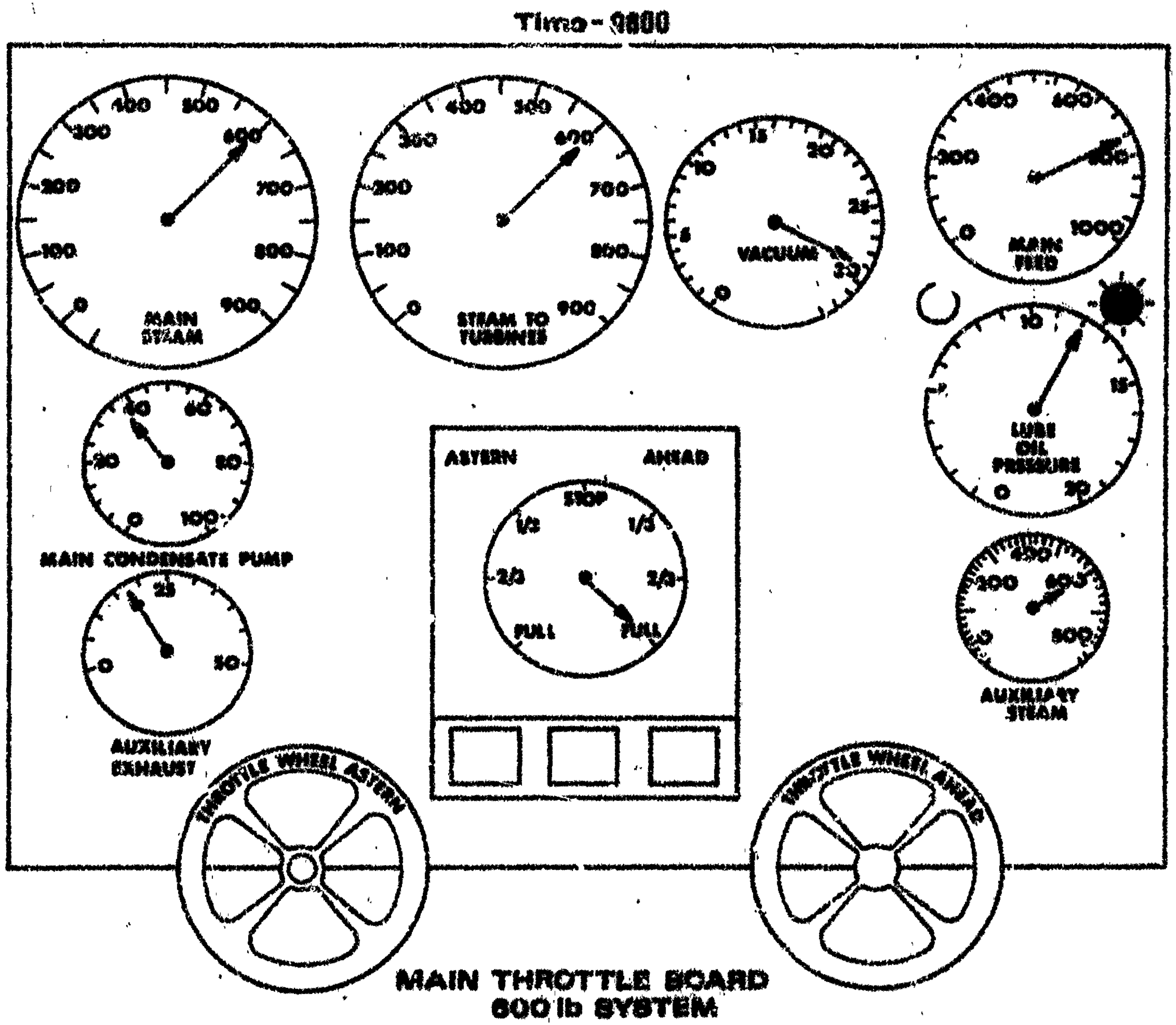

Figure 3-1. Sample throttle board test item.

Each of the four throttle board pictures consisted of eight scorable items. The examinees were given two points credit for each numerically correct answer and one point credit for each answer that was within plus or minus five per cent of the correct answer. Thus, the maximum possible score on this test was 64. No credit was given for any answer outside the five per cent limits.

Figure 3-2 shows the administration of the messenger watch test during the criterion data collection.

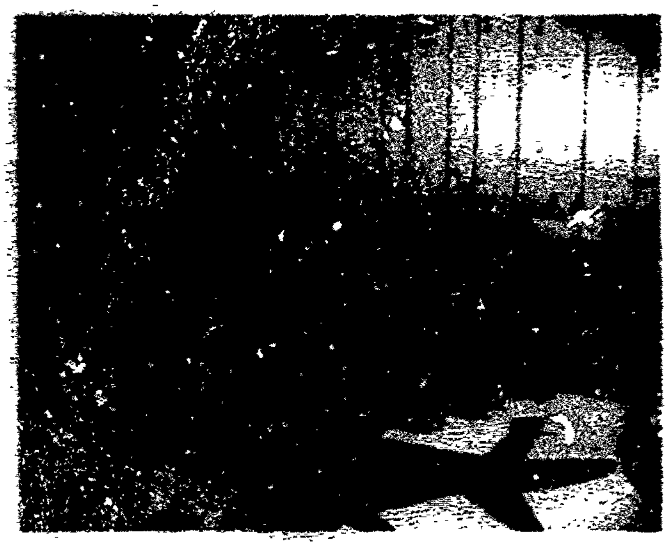

Figure 3-2. Messenger watch test situation.

Breaking-Making a Flange

The level of each recruit in the followup on breaking-making a flange was measured through an individually administered performance test. In this test, the examinee was required to break and make a flange using the following tools and items: (a) several gaskets, (b) an assembled six inch flange with input valves, (c) one scraper, (d) two combination (boxed/open end) wrenches, (e) one rag, and (f) a bucket of water with a funnel. Figure 3-3 shows one of the examinees taking the breaking-making a flange test. Scoring was accomplished through the performance checklist format.

Figure 3-3. Breaking-making a flange test situation.

Twenty scorable items were included in this checklist. Scoring was based on following the correct procedures, care and use of tools, and adherence to safety precautions. The directions to the examinees for the breaking-making a flange performance test are shown below:

"THIS IS A TEST OF YOUR ABILITY TO BREAK AND MAKE A FAULTY FLANGE. THIS IS A JOB OFTEN DONE BY MACHINIST MATES. YOUR TASK IS TO BREAK AND MAKE THIS FLANGE IN THE SAME WAY AS YOU WOULD ABOARD SHIP. AFTER YOU ARE TOLD TO BEGIN, YOU WILL HAVE 8 MINUTES TO BREAK AND MAKE THE FAULTY FLANGE. YOU WILL BE SCORED ON HOW CORRECTLY YOU PERFORM EACH STEP. YOU WILL ONLY RECEIVE CREDIT FOR THOSE ITEMS YOU FINISH CORRECTLY. ALL THE MATERIALS AND TOOLS YOU NEED ARE AVAILABLE IN FRONT OF YOU. IF YOU HAVE ANY QUESTIONS, PLEASE ASK THEM NOW."

The correct procedure, in slightly condensed form, for breaking-making a flange includes the following steps:

1. tightly closing valves on input and output sides of flange
2. pulling drainplug to drain system
3. using bucket to catch runoff from system and wiping up any spillage that occurs
4. replacing drain plug
5. braking flange apart using two boxend wrenches
6. removing old gasket
7. scraping and cleaning flange surface
8. inspection of flange surfaces for nicks, scratches or gasket residue
9. replacing correct gasket
10. replacing bolts (hand tighten)
11. tighten bolts using cross over pattern with appropriate (box-end/open end) wrenches
12. open both valves and refill system with water
13. inspection of flange for leakage

Packing a Valve

Ability on the valve packing criterion was also measured through an individually administered performance test. In this test, the examinee was required to pack a valve using: (a) a large, mounted valve, (b) packing material, (c) a knife, (d) a box end or open end wrench, and (e) a packing puller. A photograph of the tools used by the examinee for the valve packing test and the previous performance test, breaking-making a flange, is shown in Figure 3-4. Figure 3-5 shows one of the test administrators showing the available tools to an examinee. Figure 3-6 presents the examiner administering the valve packing test to an examinee.

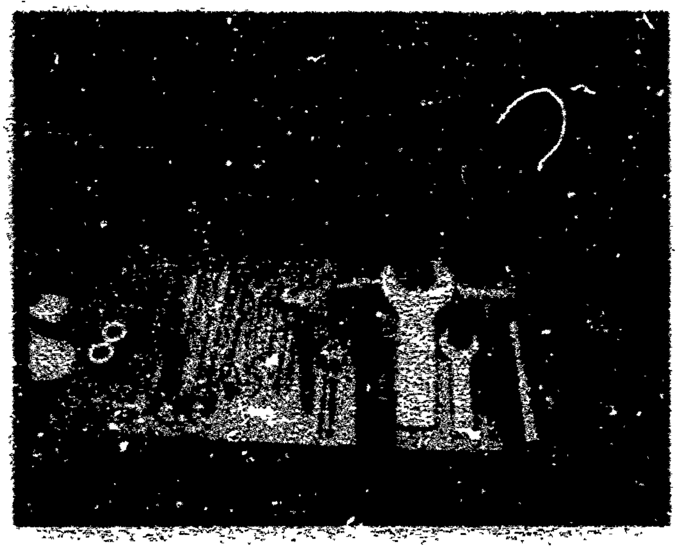

Figure 3-4. Tools used by the examinee.

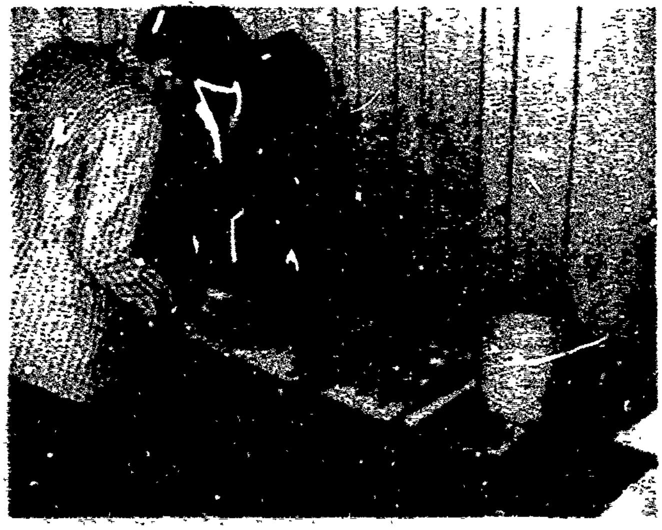

Figure 3-5. Proctor showing available tools to an examinee.

Figure 3-6. Valve packing test situation.

We note (Figure 3-4) that the examinees were given tools other than those which were needed for completing the test. Accordingly, the examinee was required to select from among the available tools those which were most appropriate for accomplishing the task. For example, the examinee had at his disposal the correct size combination wrench for loosening the nut holding the hand wheel to the valve. A crescent wrench and a monkey wrench, both of which can be adjusted to fit the handwheel nut, were also available even though they are inappropriate wrenches. Credit was lost if the improper tool was employed for a given task.

The examinee directions for the valve packing test are presented below:

"THIS IS A TEST OF YOUR ABILITY TO PACK A VALVE. THIS JOB IS OFTEN DONE BY MACHINIST MATES. YOUR TASK IS TO PACK THIS VALVE IN THE SAME WAY AS YOU WOULD ABOARD SHIP. AFTER YOU ARE TOLD TO BEGIN, YOU WILL HAVE 7 MINUTES TO PACK THE VALVE. YOU WILL BE SCORED ON HOW CORRECTLY YOU PERFORM EACH STEP. YOU WILL ONLY RECEIVE CREDIT FOR THOSE PARTS OF THE JOB YOU FINISH. ALL THE MATERIALS AND TOOLS YOU NEED ARE AVAILABLE IN FRONT OF YOU. IF YOU HAVE ANY QUESTIONS PLEASE ASK THEM NOW."

Scoring of the valve packing test was also of the performance checklist type. Twenty-three scorable items were included in this measure. Scoring was based on following the correct procedure, care and use of tools, and adherence to safety regulations. The correct procedure, in slightly condensed form, for packing a valve included the following steps:

1. removing hand wheel nut with box end or open end wrench
2. removing hand wheel
3. removing packing nut with open end wrench
4. removing packing gland
5. removing packing with packing puller
6. inspecting to see that all packing is removed
7. cut packing rings with knife
8. installing packing rings neatly and pushing down with gland until one-half of gland is showing
9. replacing gland and packing nut (hand tight)
10. replacing handwheel and handwheel nut (snugged up but not over tightened)
11. tightening with appropriate wrenches

Malfunction and Emergency Procedures (Sequential)

Knowledge of correct actions in common malfunction and emergency situations was tested through an individually administered test. Each item in the test consisted of a set of pictures depicting a Fleet emergency or malfunction correction sequence. The task of the examinee was to place the pictures, which were presented in scrambled order, in the correct sequence. To do this the examinee must first recognize what is being represented. The examinee directions for this test were:

"THIS IS A TEST OF YOUR ABILITY TO ARRANGE A SET OF PICTURES, SHOWING JOBS OFTEN DONE BY MACHINIST MATES IN THE RIGHT ORDER. WE WILL GIVE YOU A SERIES OF PICTURES. YOU ARE TO ARRANGE EACH SET OF PICTURES SO THAT THEY MAKE A SENSIBLE STORY. YOU WILL BE SCORED ON HOW WELL YOU PUT THESE PICTURES IN THE CORRECT ORDER. YOU WILL BE ALLOWED TWO MINUTES TO ARRANGE EACH SET OF CARDS IN THE CORRECT ORDER--SO THAT THEY TELL A STORY. YOU WILL BE GIVEN EXTRA CREDIT IF YOU CORRECTLY FINISH A STORY WITHIN 30 SECONDS. BEFORE WE BEGIN THE TEST LET'S DO A SAMPLE."

The various problems depicted (in order of difficulty) in the picture arrangement test were: (a) electric shock (sample item), (b) fire in compartment, (c) leaky valve repair, (d) tank gauging, (e) fire hose assembly, (f) ruptured pipe, and (g) spring bearing temperature. A sample item depicting the correct sequence for the fire in compartment item, is shown in Figure 3-7.

Scoring for each item was based on the number of correct pictorial connections. For example, if the examinee's picture arrangement was: 1,2,3,6,5,4, he would receive two points, because he correctly connected picture number 1 with picture number 2, and picture number 2 with picture number 3. On the other hand, if the examinee's picture arrangement sequence was: 1,2,3,4,5,6, he would receive five points, because he correctly ordered each pair of pictures. Also, a time bonus of one correct connection per item set was given for each item set correctly completed within 30 seconds. Thirty-three possible points could be scored in this test.

Equipment/Tools Names and Use

Equipment/tools names and use criterion level was measured through an individually administered test which consisted of a series of eight cards. Each card depicted a typical machinist mate work situation along with pictures of three tools which might be employed to complete the task. For each item, the examinee was required to select, from among the three tools shown, the best one for completing the specific job depicted. The examinee received additional credit if he was also able to name the tool and tell why it is the best tool to use. A sample item is shown in Figure 3-8.

A sample examiner proctol for the items shown in Figure 3-8 is presented below:

a) Show examinee picture of a nut with a socket.
b) Say:"THIS IS A NUT WITH A SOCKET."
c) Show examinee wrench pictures for item 2.
d) Say: "WHICH WRENCH WOULD YOU USE TO FREE THE NUT?"
e) Pause 15 seconds for his answer.
f) After a response is given continue by saying: "What IS THE TOOL CALLED?"
g) Pause 10 seconds for his answer.
h) Go on to question #3 if the first part of this question was answered correctly.

In some instances an examinee would identify the incorrect tool for a particular situation. If he was able to name the tool he identified, regardless of whether or not it was the right tool to use, he received one point credit for that item. Thirty-two possible points could be achieved on this test.

General Alertness/Common Sense (what's wrong)

General alertness/common sense criterion performance was also measured through a pictorial, individually administered test. Each item consisted of a picture of a typical machinist mate work situation in which a sailor is shown doing something wrong. The task of the examinee was to detect and report what is wrong or absurd in the picture. The examinee directions for the general alertness/common sense test are shown on the following page.

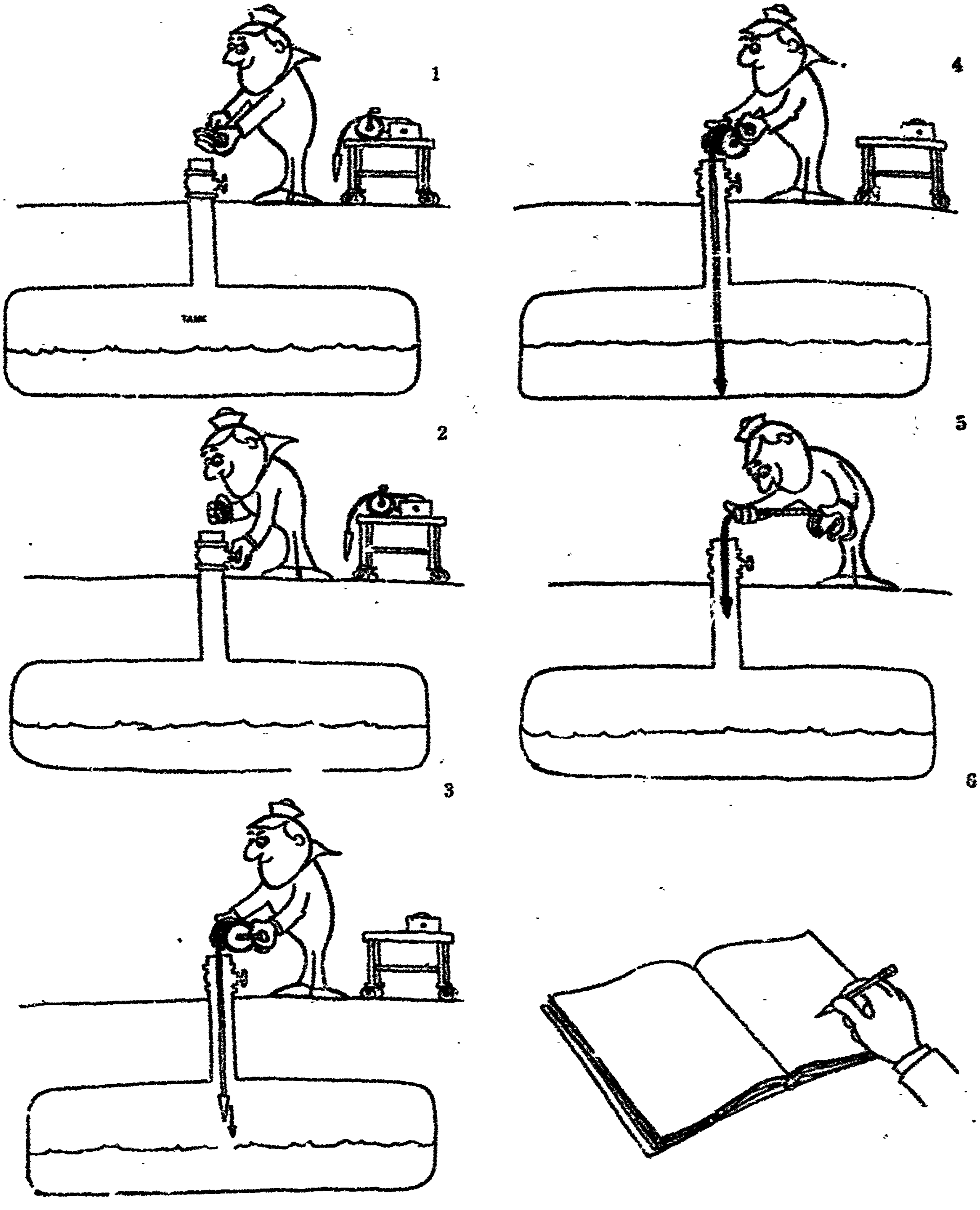

Figure 3-7. Sample malfunction and emergency procedure item
(arranged in correct sequence)

"THIS IS A TEST OF YOUR ABILITY TO SEE WHAT IS WRONG IN A DRAWING OR DIAGRAM. WE WILL PLACE THE DRAWING OR DIAGRAM IN FRONT OF YOU. YOU WILL BE ALLOWED ONE MINUTE TO TELL US WHAT IS WRONG WITH EACH DRAWING OR DIAGRAM. POINT TO THE PROBLEM ON THE DRAWING OR DIAGRAM IF YOU ARE NOT ABLE TO TELL US WHAT THE PROBLEM IS. LET'S TRY A SAMPLE PROBLEM."

Two sample examiner proctols from the general alertness/common sense test are shown below. Also shown, in Figure 3-9, is a sample general alertness/common sense test item.

1. Problem - Man using screwdriver as chisel.
   a)show examinee picture
   b)Say: "WHAT'S THE MAN IN THE PICTURE DOING WRONG?"

2. Problem - Man using chisel with a mushroom end.
   a)Show examinee picture
   b)Say: "WHAT'S THE MAN IN THE PICTURE DOING WRONG?"

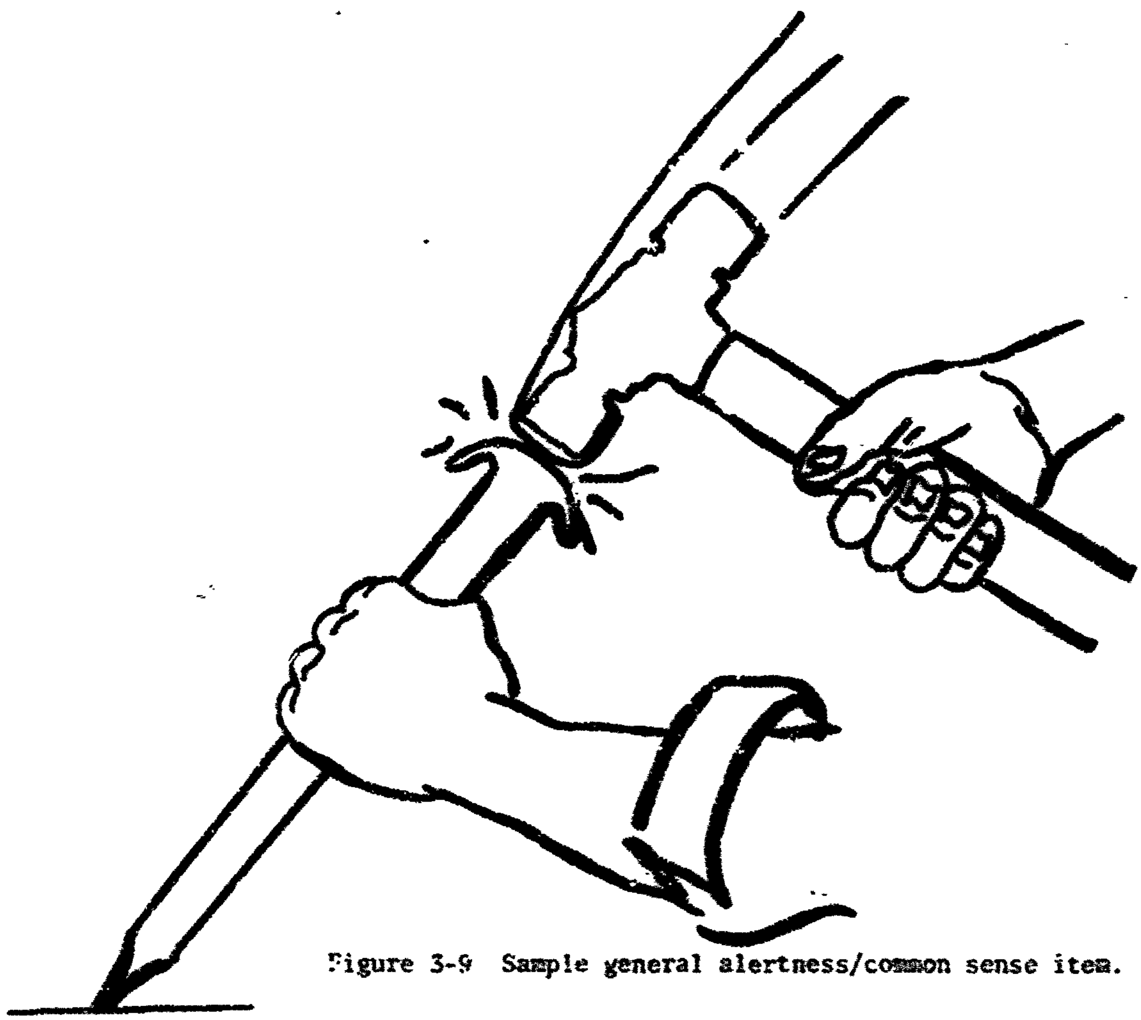

Figure 3-9 Sample general alertness/common sense item.

The items in the general alertness/common sense test were arranged in ascending order of difficulty. In scoring this test, the examinee was allowed two points for each correct answer, and one point credit if he pointed to the problem or gave a correct but not fully insightful, answer. One point was subtracted from the examinee's score for each incorrect or unanswered question. Since this test consisted of seven items, the total possible score attainable by the examinee was 14.

Establishment of Minimally Acceptable and Desirable Scores

The Delphi Technique, developed by the Rand (Helmer, 1967; Dalkey, 1967; Brown, 1968; Dalkey 1962; Dalkey, 1969; Martino, 1972) Corporation, was used to establish "minimally acceptable" and "desirable" performance scores on each of the criterion tests. The Delphi Technique is essentially, a method of converging the opinions of a small group of "experts." Each "expert" is asked a question designed to elicit a quantitative estimate or opinion. The question, in the present case, involved the "minimally acceptable" and "desirable" score (defined below) on each of the criterion tests. Each judge assigns his estimate individually without conferring with the other members of the group. In the current application, once the estimates were assigned, they were collected and placed on a blackboard so that each individual could review his estimate in the context of the other group members' estimates. Then, each "expert" was asked to justify his own opinion. Following the justification procedure, the "experts" were again asked to assign quantitative estimates to the same question. This procedure was followed for as many trials as required for the group to arrive at a convergence of opinion. Four, experienced machinist mates from the instructional staff of the Damage Control Training Center at the Philadelphia Navy Yard served as judges within the Delphi procedure.*

1. Minimally Acceptable

   The lowest score on this test that you would accept as indicating that you could make some use of this man as a machinist mate striker. This score would indicate that the man probably requires considerable supervision.

2. Desirable

   The score on this test that you would expect a man to attain before you would consider him at a desirable level of capability. This score would indicate that the man is probably able to accomplish his assigned tasks with only minimum supervision.

*One Chief Warrant Officer, one Master Chief Petty Officer, one Chief Petty Officer, and one First Class Machinist Mate.

A sample judgment for the malfunction and emergency procedures (sequential picture arrangement) test is shown in Figure 3-10.

Sequential Picture Arrangement Test

33 Total Points

__________________________ __________________________

Your name (print) Date

Directions

Please estimate what you would consider a <u>desirable</u> score for an apprentice machinist mate striker in the Fleet.

Desirable Score _______ points

Please estimate what you would consider a <u>minimally acceptable</u> score for an apprentice machinist mate striker in the Fleet.

Minimally Acceptable Score ____ points

Figure 3-10. Sample judgment form used in the Delphi procedure.

Prior to rating, the judges were given a detailed explanation of the purpose, content, and scoring of each test. Each test was reviewed item by item in order that the raters could obtain an estimate of the content and difficulty. In most cases, the test items were passed from rater to rater to allow a close examination of the items. In the cases of the Breaking-Making . Flange test and the Packing a Valve test, pictures were shown, since these tests were too bulky for presentation to the raters. In addition, the total possible score and scoring procedures (e.g., if points were subtracted for incorrect answers) were presented to the raters. At this junction the forms were completed by the raters.

In order to illustrate the dynamics of the Delphi procedure, a sample, three trial, group convergence is presented in Table 3-1.

Table 3-1

Delphi Convergence of Opinion for "Minimally Acceptable" (MA) and "Desirable" (D) Performance on the Equipment/Tools Names and Use Test

Rater	Trial 1 D	Trial 1 MA	Trial 2 D	Trial 2 MA	Trial 3 D	Trial 3 MA
1	30	25	28	22	25	19
2	24	18	26	18	26	18
3	25	19	25	19	25	19
4	24	16	27	21	24	19
$\bar{X}$	25.75	19.50	26.50	20.00	25.00	18.75

In this example, raters 2 and 3 remained steadfast in their opinions throughout all three trials. Alternatively, rater 1 decreased both of his estimates throughout the entire process, while rater 4 increased his estimate on only the "minimally acceptable" score. The final average Delphi estimates for each of the criterion tests are presented in Table 3-2.

Table 3-2

Final Delphi Estimate of "Minimally Acceptable" and "Desirable" Criterion Scores on Seven Performance Criteria

Criterion	Minimally Acceptable	Desirable
Messenger Watch Recording	39.75	51.00
Messenger Watch Malfunction Detection	4.50	9.75
Breaking-Making a Flange	8.75	12.25
Packing a Valve	9.25	13.75
Malfunction and Emergency Procedures	18.25	24.75
Equipment/Tools Names and Use	18.75	25.00
General Alertness/Common Sense	5.50	8.50

Criterion Samples

Fifty-four of the 99 recruits originally tested were available as followup subjects for the present phase of the investigation. Twenty-nine of the followup subjects were black and 25 were white. The reasons for the attrition of the sample are presented below. First, 30-35 percent of the sample left the service. It seems as though many of these individuals were behavior problems. Second, several ships on which members of the sample were assigned were deployed during the followup testing. This caused a loss of another five or six subjects. Finally, a small number of subjects seemed to be unlocatable or on sick leave.

Forty-nine of the 54 subjects were stationed and tested in the Norfolk, Virginia area.* The remaining five subjects were tested at the Charleston, South Carolina Naval Base. Approximately half of the subjects were tested in large staterooms, on messdecks, or in libraries aboard ship. The remaining subjects were tested in classrooms or conference rooms which were located in close proximity to the prior areas. There was in all cases adequate space, ventilation, and lighting for testing purposes.

A small number (6-8) of the subjects in the low aptitude sample were being utilized aboard ship in engineering ratings, (e.g., Boiler Technician, Engineman, Hull Technician) other than Machinist Mate. It was felt that these ratings were similar enough to the Machinist Mate rating that the criterion tests could appropriately be used for these individuals as well.

A trained, black test administrator administered the criterion tests to the black subjects in the sample. This black test administrator was 25 years old and college educated. In addition, he served three years in the armed forces. A 30 year old white psychologist served as test administrator for the white subjects in the followup sample.

Testing time consumed between one and one-half and two hours per examinee. Each testing session began with a short preliminary conversation in order to establish rapport with the examinee. This was followed by administration of the seven criterion measures and a subject interview. Table 3-3 shows the various number and types of ships by command on which the low aptitude subjects were stationed during their tour of duty in the Navy.

*Norfolk Naval Base, Little Creek Amphibious Base, Portsmouth Naval Shipyard.

Table 3-3

Number and Type of Ships, by Command, on Which the Low Aptitude Subjects were Stationed During Their Tour of Duty in the Navy

Command	Number	Type
SERVLANT	2	Combat Stores Ships
"	2	Oilers
"	1	Repair Ship
"	2	Stores Ships
	1	Fast Combat Support Ship
"	1	Ammunition Ship
AIRLANT	2	Aircraft Carriers
CRUDESLANT	3	Guided Missile Destroyers
"	2	" " Frigates
"	1	Escort Ship
"	1	Destroyer
SUBLANT	3	Submarine Tenders
PHIBLANT	2	Amphibious Transport Docks
"	1	Tank Landing Ship
MINEWARFARE	1	Mine Warfare Ship

Control Sample

In addition to the low aptitude criterion subjects, a sample of 27 A school graduate machinist mate strikers was also tested. This group of examinees served as a comparison and control group for the purpose of determining the average performance level of the typical machinist mate striker in the Fleet. All 27 of these control subjects were white, because the overwhelming majority of A school graduates are white. The control subjects had about the same amount of Fleet experience as the low aptitude sample. As an added control measure these control subjects were, in all cases, taken from the same ships as the low aptitude subjects. All of the control subjects were required to take the tests in exactly the same manner and under the same conditions as the low aptitude sample.

Supervisory Ratings

The immediate supervisor of each person in the low aptitude followup sample was interviewed in order to ascertain the supervisor's perception of the incumbent machinist mate striker. As mentioned earlier, we place little credence in the validity of supervisory ratings, since they are subject to a multitude of biasing influences. Nevertheless, the supervisory evaluations were collected because they have some inherent face value as a reflector of journeyman ability. The actual questions in the supervisory interview are presented in Figure 3-11.

3. Can you please list for me his strengths and other strong points?
   a.
   b.
   c.
   d.

4. Can you please list his weaknesses?
   a.
   b.
   c.
   d.

6. What are his chances of advancing to a higher rate (E_3 or E_4) during the next year? (Show card)

   ____________chances

7. When comparing him with others, at a similar level of experience, how does he perform technically? (Show card)
   a.____Better than most others at a similar level.
   b.____About the same as most others at a similar level.
   c.____Not as good as most others at a similar level.

8. When comparing him with others, at the next highest level, how does he perform technically? (Show card)
   a.____Better than most others at the next highest level.
   b.____About the same as most others at the next highest level.
   c.____Not as good as most others at the next highest level.

9. To what extent does he meet your standards of technical performance? (Show card)
   a.____He exceeds my standards of technical performance.
   b.____He meets my standards of technical performance.
   c.____He does not meet my standards of technical performance.

10. If you were given the opportunity of choosing your subordinates would you choose him? (Show card)
   a.____Yes, definitely.
   b.____Yes, if no one else were available.
   c.____No, definitely not.

Figure 3-11. Supervisory interview performance evaluation questions.

Questions 1, 2, and 5 of the supervisory interview were concerned with the proportion of time the incumbent spent performing various tasks on the job. The questions are shown in Figure 3-12.

Job Task Analysis Interviews

Within the supervisory interview, each supervisor was asked to estimate the average number of hours per week that the individual low aptitude sailor being supervised spent on various machinist mate tasks. The questions employed are presented as Figure 3-12. The purpose in collecting this information was to determine whether or not the low aptitude sample was given an opportunity to learn the machinist mate rate in the Fleet. If a significant proportion of the sample was not primarily engaged in learning the rate, then, poor criterion performance could be ascribed to factors other than low ability. The task interview questions are shown in Figure 3-12. The individual test subjects were also asked a similar set of task analytic questions. These test subject task analytic questions are presented in Figure 3-13.

1. For an average week, can you tell me how many hours per week ____________________ performs the following activities:
   Name of Subject

   a. Packing valves, making gaskets, and ensuring integrity of fittings __________ hours
   b. Standing messenger watch __________ hours
   c. Galley or kitchen duty __________ hours
   d. General clean up __________ hours
   e. Working with equipment and machinery (light off pumps, check bearing temperatures, sound tanks, lubricate machinery, etc.). __________ hours
   f. Damage control and standing firewatch __________ hours

2. What other activities, both machinist mate and non-machinist mate related, does he perform each week? And what are their hours?

   a. __________ hours
   b. __________ hours
   c. __________ hours
   d. __________ hours

5. On the average, how many hours a week of on the job training and instruction was he given during his first six months on the job?

   __________ hours

Figure 3-12. Supervisory job task analytic interview questions.

SUBJECT INTERVIEW

1. How many hours a week do you pack valves, make gaskets, and break flanges?

   __________hours

2. How many hours a week do you stand messenger watch at the throttle board?

   __________hours

3. How many hours a week do you perform galley or kitchen duty?

   __________hours

4. How many hours a week do you sweep up decks, mop up dirt and the like?

   __________hours

5. How many hours a week do you work with equipment and machinery (e.g., light off pumps, sound tanks, check bearing temperatures, lubricate machinery, etc.)?

   __________hours

6. How many hours per week are you involved in damage control activities or standing firewatch?

   __________hours

7. How many other activities do you perform each week (e.g., fixing equipment, starting up and maintaining different kinds of equipment and machinery, other non-machinist mate activities?

   a. __________hours
   b. __________hours
   c. __________hours
   d. __________hours

   How many hours do you perform these activities each week?

8. Do you think you were given a fair opportunity to learn the machinist mate rate during the past 6 months aboard ship?(Ask him to explain)

Figure 3-13. Test subject job task analytic interview questions.

CHAPTER IV

RESULTS AND DISCUSSION

Chapter IV presents the results of the Fleet performance evaluation. The chapter contains sections on: (a) criterion analysis, (b) predictive validity, (c) comparison of low aptitude and A school graduate performance, (d) differential validity, (e) supervisory interview analysis, and (f) analysis of interviews with low aptitude sample.

Criterion Analysis

In order to ascertain the relationships that exist among the criterion measures, correlation matrices were developed. These allow determination of the extent of independence and freedom from redundancy among these variables. The uniqueness of the miniature aptitude tests had been previously established (Siegel & Bergman, 1972). Table 4-1 presents the intercorrleations among the Fleet criterion measures on the basis of the scores on these measures of the 54 low aptitude subjects. Tables 4-2 and 4-3 present the means and standard deviations for the predictors and criterion measures respectively.

Table 4-1

Pearson Product Moment Correlations between the Breaking-Making a Flange (BMF), Packing a Valve (PV), Tool Knowledge and Usage (TKU), Sequential (Seq), What's Wrong (WW), Meter Reading-Messenger Watch (MR-MW), and the Troubleshooting Messenger Watch (Tr-MW) Criterion Measures, and Supervisory Ratings (SR), for 54 Low Aptitude Machinist Mate Strikers

	Criterion Measure						
	PV	TKU	Seq	WW	MR-MW	Tr-MW	SR
BMF	.11	.30	.17	.23	.07	.12	.23
PV		.01	-.11	.10	.29	.09	-.05
TKU			.05	.23	.23	.09	.18
Seq.				.10	.04	-.11	-.09
WW					.37	.16	.14
MR-MW						.15	.30
Tr-MW							.07

Preceding page blank

Table 4-2

Means and Standard Deviations on the GCT, ARI, MECH, Equipment Use and Nomenclature (EUN), Gasket Cutting (GC), Meter Reading (MR), Troubleshooting (T), Equipment Operation (EO), and Assembly (A) Tests for 54, Low Aptitude, Machinist Mate Strikers

Test	Mean	S.D.
GCT	39.20	7.07
ARI	42.39	5.17
MECH	42.31	5.18
EUN	20.80	2.84
GC	14.67	2.66
MR	11.81	3.88
T	15.37	5.12
EO	59.72	6.31
A	23.41	3.61

Table 4-3

Means and Standard Deviations on the Breaking-Making a Flange (BMF), Packing a Valve (PV), Tool Knowledge and Usage (TKU), Sequential (Seq), What's Wrong (WW), Meter Reading-Messenger Watch (MR-MW), and Troubleshooting-Messenger Watch (Tr-MW) criterion Measures and Supervisory Ratings (SR) for 54 Low Aptitude Machinist Mate Strikers

Test	Mean	S.D.
BMF	6.41	1.84
PV	9.43	2.82
TKU	20.02	4.83
Seq	11.61	3.65
WW	3.59	3.25
MR-MW	55.91	5.42
Tr-MW	13.74	2.02
SR	-.30*	17.68*

*Based on a summation of standard scores.

Predictive Validity

Three separate types of analyses were performed in order to ascertain the relationship between the predictor instruments (miniature aptitude tests and Navy predictors) and the criterion measures. These analyses were performed primarily to assess the predictive power of the miniature aptitude tests and secondarily to compare the predictive power of the miniature aptitude tests with the predictive power of the usual Navy tests.

In the first analysis, composite predictor scores were correlated with a composite criterion measure for both the A school graduate sample and the low aptitude sample. The specific predictor composite scores involved were: (1) composite Navy test scores consisting of the sum of the scores on the three Navy predictor tests, and (2) the sum of the standard scores of the six miniature aptitude tests. The composite field criterion measure consisted of the sum of standard scores of the Fleet performance tests. These composite score correlations were calculated in order to satisfy the economic and logical aspects inherent to any validity study, i.e., to provide a useful, summary index of the predictor-criterion relationship. A single correlation measure as opposed to a series of correlations can be most useful in some descriptive circumstances. The Pearson product moment correlation coefficient between the composite Navy predictors and the composite field criterion for the A school graduates was -.01. This indicates, at least in the context involved, that the Navy predictors possessed no predictive relationship with the criterion employed.

The product moment correlation coefficients between the composite Navy and the miniature aptitude predictors and the composite field criterion for the low aptitude sample were .33 and .49 respectively. While there is no statistically significant difference between these coefficients, the miniature aptitude predictors accounted for more than twice as much predictable criterion variance as did the Navy predictors. It has been suggested that the usual Navy test-composite criterion correlation should be corrected for the restricted range of low aptitude sample on the usual Navy tests. We do not believe that a correction for restriction in range is appropriate here because the range of Navy test scores in the low aptitude population is, in fact, restricted. The findings indicate that if the low* and high aptitude data are combined so as to eliminate range restriction, the composite predictor-criterion correlation for the Navy predictors increase to a value of .53. This grouping, though, leads to the spurious conclusion that the Navy tests are equally as powerful as the miniature aptitude tests ($r = .49$) for predicting Fleet criterion performance for the low aptitude sample. This conclusion is held to be spurious on both logical and statistical grounds.

On a logical basis, it would be incorrect to combine the high and low aptitude groups, because the subgroups comprising this combination differ qualitatively from each other. These qualitative differences consist of factors such as educational opportunity, reading habits, racial composition, interests, and motivation. In a sense, combining the low and high aptitude groups in the present investigation is akin to performing a validity study in which clerical workers and bus drivers are combined.

*A random sample of 27 low aptitude subjects was combined with the sample of 27 high aptitude subjects.

When combining the two distict groups, each of which exhibits either moderate or negligible predictor-criterion correlations, it is possible to increase spuriously the correlation if the following condition is met. If the high aptitude group, has both higher predictor and criterion scores than the low aptitude group, then combining these two groups will result in a spuriously high correlation. Figure 4-1 graphically illustrates this point. Each of the solid plots in this figure represent the high and low aptitude predictor-criterion scatterplot of scores. The broken ellipse in the figure presents the combined scatterplot. It is clear from the figure, that the combined ellipse yields much greater predictability than either of the two separate scatterplots.

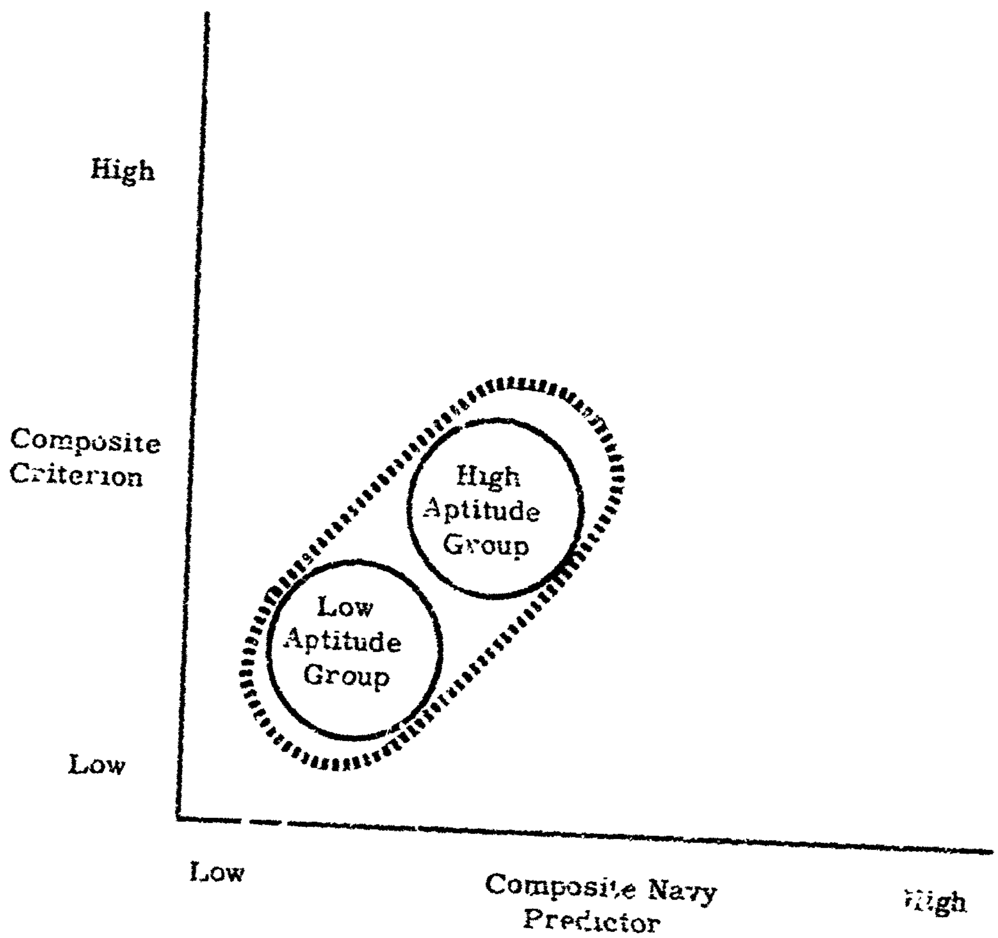

Figure 4-1. Plot of spurious increase in correlation through data combination.

Finally, if the data were available the composite miniature aptitude predictor-composite criterion correlation could also be spuriously elevated. We have already observed that the A school subjects when considered collectively, score higher on the composite criterion than the low aptitude subjects. In addition, it is exceedingly likely that the A school subjects would also score somewhat higher on the composite miniature aptitude predictors. Unfortunately, these latter data are not available. We estimate though that this correlation would increase to at least .65 if the data were available.

The second set of correlational analyses involved the calculation of the multiple correlation coefficient between: (1) the Navy predictors and the composite criterion, and (2) the miniature aptitude predictors and the composite criterion. The multiple correlation coefficient between the three Navy predictors and the composite criterion was found to be .36. The multiple correlation between the six miniature aptitude predictors and the composite criterion measure was found to be .50. The results are directionally analogous to those derived from the simpler composite predictor and composite criterion case. When only three of the miniature aptitude tests were included as predictors, the correlation was still relatively high--.46. Again, the miniature aptitude tests account for twice as much behavioral variation as the usual Navy predictor tests.

In the final correlation analysis, multiple correlation coefficients were calculated between the two sets of predictors (Navy and miniature aptitude) and each Fleet criterion measure. These data analyses allow deeper insight into the relative contribution of various predictors to the overall correlation and also provide a weighted basis for prediction, on an individual criterion basis, of criterion score from predictor scores.

Table 4-4 shows the multiple correlation coefficients for the prediction of field criterion performance by the Navy and the miniature aptitude predictors. Table 4-4 indicates that the GCT, ARI, and MECH composite predicted with statistical significance ($p < .05$) the tool knowledge and usage, meter reading-messenger watch, and what's wrong criteria for the 54 low aptitude Navy machinist mate strikers while the miniature aptitude tests predicted performance ($p < .05$) on the supervisory ratings, tool knowledge and usage, what's wrong, and meter reading-messenger watch criteria. Here again, there is a tendency for the miniature aptitude tests to be more predictive than the usual Navy tests. Not only was one more statistically significant correlation coefficient yielded but also of the eight pairs of coefficients included in Table 4-4, the value for the miniature aptitude tests is greater than that for the usual Navy tests in six cases. The directional difference, as tested by the "sign test," is statistically significant at the .145 level of confidence.

Table 4-4

Multiple Correlation between the Two Sets of Predictors* and the Individual Criterion Measure (N= 54)

Criterion	Predictor	
	Navy	Miniature Aptitude
Breaking-Making a Flange	.17	.30
Packing a Valve	.14	.22
Tool Knowledge and Usage	.47**	.42**
Sequential	.33	.35
What's Wrong	.39**	.43**
Meter Reading-Messenger Watch	.27	.46***
Troubleshooting-Messenger Watch	.43**	.22
Supervisory Ratings	.32	.39**

*Only the best three miniature aptitude predictors were utilized in order that direct comparisons between the two types of predictors could more readily be made.
** $p<.05$
*** $p<.01$

A complete listing of the multiple regression equations for predicting the individual criteria from the various predictors is presented in the appendices of this report.

Comparison of Low Aptitude and A School Graduate Performance

Another question of concern is that of whether or not recruits who show little promise of succeeding in the machinist mate rate, according to the traditional Navy tests (GCT, ARI and MECH), can be identified by other predictors (miniature aptitude tests) and, in fact, be shown to succeed according to an "on the job" field criterion. To this end t-tests were performed to ascertain the statistical significance of the performance differences, if any, between the 36 low aptitude subjects who were predicted to be successful in the Fleet* and the 27 A school graduates who met the Navy aptitude test requirements. The mean and the standard deviation of the low aptitude subjects who were judged to "pass" the miniature aptitude screen and the A school graduate samples on each criterion test are shown in Table 4-5. According to Table 4-5, the performance of the A school graduates was significantly superior to that of the low aptitude sample (those who failed on traditional Navy aptitude tests) on five of the seven field criterion tests.

*A subject was predicted to be successful in the Fleet if he passed the troubleshooting test and any of two of the remaining five tests in the miniature aptitude battery.

Table 4-6 presents the percentage of low aptitude subjects predicted to achieve some degree of field success, according to miniature aptitude predictors, who attained or exceeded the 25th and 50th percentile level of scores achieved by the A school graduates on the field criterion tests.

Clearly, a considerable number of the members of the low aptitude group, in many cases, performed on these criterion tests as well as, or even better than, some of the high aptitude A school graduates.

Table 4-7 shows the performance of the A school graduate and the low aptitude groups on the basis of the percentage of individuals from each group whose criterion scores fell into each of the three criterion referenced categories--desirable, minimally acceptable, and below minimally acceptable--as derived from the Delphi application. On the whole, more of the criterion scores of the A school graduates fell into the upper two categories. However, the data of Table 4-7 also indicated that: (1) eight per cent of the low aptitude machinist mates performed at or above the minimally acceptable level on the Breaking-Making a Flange test, (2) 61 per cent performed above the minimally acceptable level on the Packing a Valve test, (3) 79 per cent performed above the minimally acceptable level on the Tool Knowledge and Usage test, (4) 6 per cent fell above the minimally acceptable level on the Sequential test, and (5) 28 per cent performed above the minimally acceptable level on the What's Wrong test, (6) 89 and 100 per cent respectively scored above the minimally acceptable level on the Meter Reading and Trouble Shooting tests. Accordingly, it seems that the miniature aptitude tests can glean from a group of non Navy test qualified persons those individuals who can reach Fleet performance criteria with six months Fleet experience. This could represent a considerable harvest in terms of both manpower and equity. Vineberg and his colleagues (Vineberg, Taylor, & Caylor, 1970; Vineberg & Taylor, 1972a; Vineberg & Taylor, 1972b) have also found that a substantial portion of low aptitude subjects can perform adequately on military jobs.

Differential Validity

Given the findings in Table 4-7, the questions of whether or not the miniature aptitude tests are more "culture-fair" than the traditional Navy predictors, and, more specifically, whether the miniature aptitude tests are differentially valid remain to be answered.

The first step in determining the answer to these questions was to determine if the composite Navy and miniature aptitude predictor scores were significantly different across the 29 black and 25 white low aptitude subjects. T-tests were performed in order to ascertain if the scores on the predictors did, indeed, exhibit statistically significant differences between races.

No statistically significant differences were found between the scores for the black and the white samples on both the Naby predictors and the miniature aptitude tests ($t = 0.94$ and 1.06 respectively, $p > .05$). It was found, however, that the performance on the composite

Table 4-5

Performance of A School Graduate (ASG) and Low Aptitude (LA) Navy Machinist Mate Strikers on Field Criterion Tests

	Breaking and Making a Flange		Packing a Valve		Tool Knowledge and Usage		Sequential		What's Wrong		Meter Reading Messenger Watch		Troubleshooting Messenger Watch	
	ASG	LA	ASG	LA	ASG	LA	ASG	LA	ASG	LA	ASG	LA	ASG	LA
$\bar{x}$	7.64	6.42	10.11	9.64	26.44	20.22	13.93	11.58	9.22	7.44	59.89	56.97	14.78	13.72
σ	1.60	1.76	2.79	2.86	2.92	4.97	5.53	3.77	2.83	2.90	1.80	5.35	1.53	20.4
t	2.85**		.657		6.21**		1.89		2.44*		3.05**		2.54*	

*Significant between the .05 and the .01 levels of confidence.

**Significant at or below the .01 level of confidence.

Table 4-6

Percentage of Low Aptitude Machinist Mate Strikers Predicted by the Miniature Aptitude Test Predictors to Pass Who Attained or Exceeded the 25th and 50th Percentile Levels of the A School Graduates

	Breaking and Making a Flange	Packing a Valve	Tool Knowledge and Usage	Sequential	What's Wrong	Meter Reading Messenger Watch	Troubleshooting Messenger Watch
50th percentile	25	47	11	30	30	44	50
25th percentile	66	61	27	47	32	58	58

Table 4-7

Percentage of Machinist Mate Strikers in the A School (N= 27) and in Predicted Successful Low Aptitude Sample (N= 36) in the Three Delphi Categories for Each Field Criterion Test

	Delphi Level	BMF	PV	TKU	Seq.	WW	MR	TR
	Desirable	4	7	67	4	22	100	100
A	Minimally Acceptable	18	71	33	18	41		
School	Below Minimally Acceptable	78	22		78	27		
	Desirable		3	32		8		100
Low	Minimally Acceptable	8	58	47	6	20	89	
Aptitude	Below Minimally Acceptable	92	39	31	94	72	11	

field criterion data did approach a statistically significant difference (t= 1.98) in favor of the whites. Examination of these data indicate that the difference is the result of an accumulation of small differences in each of a number of the performance criteria. It is possible that differential job exposure was afforded the whites as compared to the blacks and that this differential exposure contributed to the differences noted. To provide additional insight into the differential validity question, the relationships among the composite miniature aptitude and Navy predictors and the composite field criterion for both blacks and whites were examined. Table 4-8 presents the composite predictor-composite criterion product moment correlation coefficients for both black and the white low aptitude subjects. Table 4-8 indicates that both the Navy and the miniature aptitude predictor-criterion correlation coefficients failed to achieve desirable predictive levels for whites in terms of the composite field criterion. However, both the miniature aptitude and the Navy tests predicted the performance of the black low aptitude machinist mate strikers on the composite criterion. Differential validity is said to obtain (Boehm, 1971) when the correlation coefficients for two groups differ significantly from zero and from each other. The correlation coefficients presented in Table 4-8 were converted to z scores and tests of the statistical significance of the difference between the correlation coefficients for the black and the white low aptitude groups were completed. None of the differences was found to approach .05 level of confidence. Accordingly, both of the predictor batteries fail to meet Boehm's criteria for differential validity and can not be held to be differentially valid.

Table 4-8

Pearson Product Moment Correlation Coefficients (by Race) Between Composite Criterion and Composite Predictor Scores

Race		Predictor: Navy	Predictor: Mini. Apt.
Black	r	.39*	.60*
	y int.	43.4	16.8
White	r	.23	.23
	y int.	74.7	74.7

*Statistically significant at or below the .05 level of confidence

Supervisory Interview Analysis

The results of the supervisory interview analysis indicated that a large proportion (actually a majority) of the low aptitude subjects to have performed, at least, adequately (in the opinion of their immediate supervisor) in their first six months of service. In many instances, the subjects performed at a better than adequate level.

When the supervisors were asked to compare the performance of their low aptitude subordinates with the performance of others, at a similar level of experience, 29 per cent were judged as performing better than most others at a similar level. Fifty per cent were said to be performing about the same as most others at a similar level, and 21 per cent were said to be performing poorer than most others at a similar level. Seventy-nine per cent of the low aptitude subjects, then, seem to be performing at least at about the same criterion level as most others with similar experience. When asked to compare the low aptitude subjects with others at the next highest level, ten per cent were judged as performing better than most others at the next highest level while 41 per cent of the subjects were judged as performing the same as those at the next highest level. Forty-eight per cent were performing less well than those at the next highest level. Roughly half of the low aptitude subjects, then, were performing as well as those at the next highest level.

According to the supervisors, 14 per cent of the low aptitude subjects exceeded the standards of technical performance of the supervisors. Sixty-five per cent of the subjects met their supervisors' standards while only 19 per cent failed to meet those standards. Eighty-one per cent of the low aptitude sample, then, appear to be meeting supervisory performance standards.

Finally, 62 per cent of the supervisors reported that they would choose their particular low aptitude subordinate if given the opportunity of choosing subordinates. Twenty-nine per cent of the time they indicated they would choose him if no one else were available. In only nine per cent of the cases would they definitely not choose him as a subordinate.

The supervisors of the low aptitude subjects were also asked to list the strengths of the low aptitude sample. Table 4-9 lists the number of times various strengths were mentioned and the proportion of the low aptitude sample to whom each descriptor was applied.

Many of the strengths in Table 4-9 are not those which one would ordinarily associate with a low aptitude group. Yet, they do support the previous data regarding the positive performance of a number of the low aptitude subjects.

Table 4-9

Strengths* of Low Aptitude Navy Machinist Mate Strikers

Strength	Number	Proportion**
Motivation, Initiative	37	64
Intelligent, Quick Learner	15	26
Obeys Orders	14	24
Does good job, Adequate job	11	19
Dependable, Reliable	9	16
Gets along with others	8	14
Mechanically inclined	6	10
None	5	8
Learned job	4	7
Good Behavior	3	5
Good Morale	3	5

*Strengths and strong points mentioned only once were not included in this table. The N for this table is 58 not 54.

** The supervisors of four low aptitude subjects were available even though the subjects themselves were not available for criterion testing. The sum of the proportions are greater than 100, inasmuch as some supervisors mentioned more than one strength or string point per subject.

Table 4-10 lists the weaknesses of the low aptitude subjects, as indicated by their supervisors. Table 4-10 indicates that poor learning ability and lack of motivation were the primary problem areas exhibited by the low aptitude subjects. Other problems mentioned with intermediate frequency are need for supervision and discipline.

Table 4-10

Weaknesses* of Low Aptitude Navy
Machinist Mate Strikers

Weakness	Number	Proportion**
Poor learner, Slow learner, forgetful	16	28
No motivation, No initiative, Lazy	15	26
Needs supervision	10	17
None	9	16
Discipline	8	14
Verbal and Written Material	5	8
Withdrawn	2	4
Untrainable	2	4

*Weaknesses mentioned only once were not included in this table. The N for this table is 58 not 54. The supervisors of four low aptitude subjects were available even though the subjects themselves were not available for criterion testing.

** The proportions are greater than 100, inasmuch as some supervisors mentioned more than one strength or strong point per subject.

Low Aptitude Interview Analysis

Thirty-nine per cent of the tested low aptitude subjects indicated that they thought that they were not given a fair opportunity to learn the machinist mate rate. The chief reasons for this opinion, as related by the respondents, were that:

1. they were not properly taught
2. they were ignored
3. they were assigned menial tasks
4. they were on galley duty

Some of the other interview questions provide additional insight into the reasons for the perception on the part of the low aptitude sample that they were not given a fair opportunity to learn the rate. Forty-four per cent of the sample reported that they had served over three months on galley duty. In addition, 33 percent indicated that they had little opportunity to pack valves, break flanges, or make gaskets. Evidently, the request included with the assignment of each low aptitude subject to a ship, that he be assigned to work within the machinist mate specialty, was often ignored. The additional request, that the subject be given every opportunity to learn the tasks typical of the rate was also ignored in many instances. The implication to be drawn from these data, then, is that a large proportion of the low aptitude sample was not given the opportunity to perform in or learn the machinist mate rate. This fact probably serves to introduce a random factor into our predictor-criterion relationships.* Accordingly, the correlations reported here are, in all likelihood, underestimates of true predictor-criterion relationships.

From the Navy supervisor point of view, one could say that a large proportion of the low aptitude subjects are slow learners. The supervisors may not have had the time or the knowledge/skills to administer the special types of training required for slow learners. Hence, the slow learner was placed in work situations requiring little or no training (e.g., menial tasks).

Summary and Conclusions

This research program was based on a testing concept related to "culture fairness" and specifically to an interpretation which contends that if an individual can learn to perform a job sample, he can also learn to perform the total job. The demonstrated ability to learn selected aspects of a job is employed as a predictor of ability to learn to perform the total job.

The machinist mate rate was selected as a logical first rate to do initial research with the innovative testing concept. The machinist mate rate involves performance of tasks which are largely nonverbal in nature, and thus would allow an individual who, for whatever reason, lacks verbal-ability to excel.

A set of job-related miniature aptitude tests was constructed and administered to Navy recruits who had failed on traditional Navy predictors (presumably because these tests are "loaded" with verbally culture-related materials) and thus were not eligible for machinist mate training in the A school. The job sample involved no written learning materials; hence, the job sample maximally simulated on-the-job training situation in which a foreman instructs a journeyman in job performance. After testing, the recruits were placed aboard ship for assignment as an entry level machinist mate. Criterion data were collected after the recruits had six months of Fleet experience. Out of a total of 99 low aptitude recruits tested with the miniature aptitude job sample tests, performance criterion data were available to be collected for a sample of 54. Criterion data were also collected on 2? A school graduate machinist mate strikers. The high aptitude machinist mate strikers were found, on the whole to perform significantly better on the performance criterion shipboard tests than did the low aptitude machinist mates. However, members of the low aptitude group did, in many cases perform on these job sample performance

*Note the low relationships found between the predictors and the flange making and valve packing criterion tests.

criterion tests as well as, or even better than, some of the high aptitude A school graduates. The data also indicated that the miniature aptitude test predictors correlated higher than the usual Navy predictors with the field criteria tests for the low aptitude sample, and that the miniature aptitude tests accounted for more than twice the criterion variance as compared with the Navy predictors.

The results of interviews with the immediate work supervisors of the low aptitude sample indicated that a large proportion were performing at an acceptable level. Specifically, 81 per cent of the low aptitude sample were considered by their supervisors to be performing at an acceptable level. The areas most frequently mentioned by the supervisors as descriptive of the strengths of the low aptitude sample were: "motivation-initiative," "quick learner," and "obeys orders." The most frequently mentioned weak areas of the "low aptitude" sample were: "poor learner," "no initiative," and "needs supervision."

The data did not support a conclusion that the miniature aptitude tests are differentially valid. Accordingly, their usefulness for both black and white low aptitude machinist mate applicants seems supported, at least tentatively.

The primary goal of the present study--developing a culture fair technique that can validly predict performance of low aptitude Navy machinist mate applicants--has, at least partially, been achieved. It is thus suggested that tests of this sort represent a useful tool in regard to man-power and equity. Similar research with similar tests but with other Navy rates would seem to be an appropriate pursuit in the future.

The next phase of the present program involves following as many of the initial sample for a second set of criterion tests. This second set of criterion tests will be administered after the sample has had about one year of Fleet experience.

REFERENCES

Abramson, E. Levels of aspiration of Negro 9th. grade males in integrated and segregated schools. *Psychological Reports*, 1971, *29(1)*, 258.

Bartlett, D., Newbrough, J., & Tulkin, S. Raven Progressive Matrices: An item and set analysis of subjects grouped by race, sex, and social class. *Journal of Consulting and Clinical Psychology*, 1972, *38(1)*, 154.

Brown, B. *Delphi process: A methodology used for the elicitation of opinions of experts.* Santa Monica: Rand Corp., 1968.

Bucky, S. & Banta T., Racial factors in test performance. *Developmental Psychology*, 1972, *6(1)*, 7-13.

Chappell, R., Brenner, M., Dammond, D., Harrison, M., Lee, C., & Stone, R. The programs of seven companies to utilize the hard-core unemployed. *National conference of the effective utilization by industry of the hard-core unemployed*, May 2, 1968.

Clark, K. Some problems in the adaptation of ghetto youth to industry. *National conference on the effective utilization by industry of the hard-core unemployed*. May 2, 1968.

Dalkey, N. An experimental study of group opinion. *Futures*,1969, *2*.

Dalkey, N. *Delphi*. Santa Monica: Rand Corp., 1967.

Dalkey, N. & Helmer, O. *The use of experts for the estimation of bombing requirements: A project Delphi experiment.* Santa Monica: Rand Corp., 1962.

Flaugher, R., & Rock, D. Patterns of ability factors among four ethnic groups. *Proceedings of the Annual Convention of the American Psychological Association*,1972, *7(1)*, 27-28.

Gael, S., & Grant, D. Employment test validation for minority and non-minority telephone company service representatives. *Journal of Applied Psychology*, 1972 , *56(2)*, 135-139.

Garcia, J. I.Q.,the conspiracy. *Psychology Today*, , *6(4)*,40-43.

Getsinger, S., Kunce, J., Miller, D., & Weinberg, S. Self-esteem measures and cultural disadvantagement. *Journal of Consulting and Clinical Psychology*, 1972 , *38(1)*, 149.

Gould, L., & Klein, E. Performance of black and white adolescents on intellectual and attitudinal measures as a function of race of tester. *Journal of Consulting and Clinical Psychology*, 1971, *37(2)*, 195-200.

Preceding page blank

Green, R. On the correlation between I.Q. and amount of "white" blood. *Proceedings of the Annual Convention of the American Psychological Association*, 1972, *7(1)*, 285-286.

Green, R. & Rohwer, W. SES differences on learning and ability tests in black children. *American Educational Research Journal*, 1971, *8(4)*, 601-609.

Hallahan, D. Cognitive styles: Preschool implications for the disadvantaged. *Journal of Learning Disabilities*, 1970, *3(1)*, 4-9.

Helmer, O. *Analysis of the future: The Delphi method.* Santa Monica: Rand Corp., 1967.

Helmer, O. *Systematic use of expert opinions.* Santa Monica: Rand Corp., 1967.

Humphreys, L. Implications of group differences for test interpretation. In *Assessment in a pluralistic society.* Princeton: Educational Testing Service, Pp. 56-71.

Hunt, J. Heredity, environment and class or ethnic differences. In *Assessment in a pluralistic society.* Princeton: Educational Testing Service, Pp. 3-36.

Katz, I. A critique of personality approaches to negro performance with research suggestions. *Journal of Social Issues*, 1969 , *15(3)*, 13-27.

Lefebre, A. The relationship between self-concept and level of aspiration with negro and white children. *Dissertation Abstracts*, 1971 , *32(5-B)*, 3036.

Martino, J. The precision of Delphi estimates. *Technological Forecasting and So... Change*, 1972 , *1(3)*, 293-299.

Mercer, J. I.Q., the lethal label. *Psychology Today*, 1972 , *6(4)*, 44-47.

Ruch, W. A re-analysis of published differential validity studies. *Symposium, differential validation under EEOC and OFCC testing and selection regulations.* Honolulu: American Psychological Association, Sept. 6, 1972.

Siegel, A. I., & Bergman B. *Nonverbal and culture fair performance prediction procedures. I. Background, test development, and initial results.* Wayne, Pa.: Applied Psychological Services, 1972.

Siegel, A.I., Bergman, B., Federman, P., & Sellman, W. *Some techniques for the evaluation of technical training courses and students.* Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, February, 1972. (AFHRL-TR-72-15).

Soares, A. & Soares, L. Comparative differences in the self perceptions of disadvantaged and advantaged students. *Journal of School Psychology*, 1971 , *9(4)*, 424-42?.

Southern, M. & Plant, W. Differential cognitive development within and between racial and ethnic groups of disadvantaged preschool and kindergarten children. *Journal of Genetic Psychology*, 1971, *119(2)*, 259-266.

SPSSI Council statement on race and intelligence. *Journal of Social Issues*, 1969 , *25(3)*, 1-4.

Tobias, R., Busby, E., Greenfield, R., Goldberg, N. & Moed, M. The techniques used by four poverty agencies to prepare the hard-core for industry. *National Conference on the effective utilization by industry of the hard-core unemployed,* May 2, 1968.

Veroff, J. & Peele, S. Initial effects of desegregation on the achievement motivation of negro elementary school children. *Journal of Social Issues,*1969, *25(3)*,71-91.

Vineberg, R., Taylor, E., & Caylor, J. *Performance in five Army jobs by men at different aptitude (AFQT) levels: 1. Purpose and design of study.* HumRRO Technical Report, 1970, 70-18.

Vineberg R., & Taylor E. *Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance.* HumRRO Technical Report, 1972, 72-22.

Vineberg, R. & Taylor E. *Performance in four Army jobs by men at different aptitude (AFQT) levels: 4. Relationships between performance criteria.* HumRRO Technical Report, 1972, 72-23.

Watson, P. I.Q. The racial gap. *Psychology Today,*1972 , *6(4)*,48-50.

Wilcox, R. Racial differences in associative style. *Language and Speech,* 1971 , *14(3)*, 251-255.

Williams, N. Race, poverty and educational achievement in an urban environment. *Proceedings of the Annual Convention of the American Psychological Association,*1972 , *7(1)*, 467-468.

Willie, C. A theoretical approach to cultural and biological differences. In, *Assessment in a pluralistic society.* Princeton: Educational Testing Service, Pp. 72-84.

Yando, R., Ziegler, E., & Gates, M. The influence of negro and white teachers rated as effective or non-effective on the performance of negro and white lower-class children. *Developmental Psychology,* 1971, *5(2)*, 290-299.

APPENDIX A

Regression Equations for Predicting Criterion Performance
from Navy Basic Battery Test Scores

KEY:

GCT = General Classification Test
ARI = Arithmetic Test
MECH = Mechanical Test
BMF = Breaking Making a Flange Test
PV = Packing a Valve Test
TKU = Tool Knowledge and Usage Test
Seq = Sequential Picture Arrangement Test
WW = What's Wrong Test
MW-MR = Messenger Watch-Meter Reading Test
MW-TR = Messenger Watch-Troubleshooting Test

Preceding page blank

Table A-1

Regression Equation for Predicting Criterion Performance on the BMF Test Using the GCT as a Predictor (N= 54)

Variable	Multiple	Regression Equation*
GCT	..06	$\hat{y} = 8.1 - .04(x_1)$

* x_1=GCT. F level insufficient for computation of the other variables.

Table A-2

Regression Equations for Predicting Criterion on Performance on the PV Test Using GCT, ARI and MECH as the Predictors (N= 54).

Variables	Multiple R	Regression Equation*
GCT	.119	$\hat{y} = 7.56 + .05(x_1)$
GCT + ARI	.129	$\hat{y} = 6.63 + .04(x_1) + .03(x_2)$
GCT + ARI + MECH	.141	$\hat{y} = 7.64 + .84(x_1) + .04(x_2) + .0[illegible](x_3)$

x_1= GCT
x_2= ARI
x_3= MECH

Table A-3

Regression Equation for Predicting Criterion Performance on the TKB using GCT, MECH and ARI as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
MECH	.468**	$\hat{y} = 1.54 + .43(x_3)$
MECH + GCT	.471**	$\hat{y} = 2.84 - .04(x_1) + .44(x_3)$
MECH + GCT + ARI	.472**	$\hat{y} = 3.59 - .03(x_1) - .03(x_2) + .45(x_3)$

* x_1 = GCT
x_2 = ARI
x_3 = MECH

Table A-4

Regression Equation for Predicting Criterion Performance on the GCT Seq. Test using GCT, ARI and MECH as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
GCT	.217	$\hat{y} = 7.21 + .11(x_1)$
GCT + MECH	.279	$\hat{y} = 2.23 + .10(x_1) + .12(x_3)$
GCT + MECH + ARI	.334	$\hat{y} = 5.39 + .15 x_1 - .15(x_2) + .15(x_3)$

* x_1 = GCT
x_2 = ARI
x_3 = MECH

Table A-5

Regression Equations for Predicting Criterion Performance on the WW Test Using GCT, ARI and MECH as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
GCT	.307	$\hat{y} = .14(x_1) - 2.16$
GCT + MECH	.379*	$\hat{y} = .13(x_1) + .14 - 7.74$
GCT + MECH + ARI	.385*	$\hat{y} = .15(x_1) - .05 + .15(x_3) - 6.74$

*x_1= GCT
x_2= ARI
x_3= MECH

Table A-6

Regression Equations for Predicting Criterion Performance on the Messenger Watch-Meter Reading Test Using the GCT, ARI and MECH as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
MECH	.201	$\hat{y} = 46.99 + .21(x_3)$
MECH + GCT	.254	$\hat{y} = 42.87 + .12(x_1) + (x_3)$
MECH + GCT + ARI	.271	$\hat{y} = 40.43 + .09(x_1) + .11(x_2) + .17(x_3)$

*x_1= GCT
x_2= ARI
x_3= MECH

Table A-7

Regression Equations for Predicting Criterion Performance on the Messenger Watch-Troubleshooting Test Using the GCT, ARI and MECH as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
ARI	.371**	$y = 6.84 + .16(x_3)$
ARI + MECH	.427*	$y = 8.03 + .17(x_2) - .04(x_3)$
ARI + MECH + GCT	.434*	$y = 8.44 - .02(x_1) + .19(x_2) - .04(x_3)$

*x_1 = GCT
x_2 = ARI
x_3 = MECH

Table A-8

Regression Equations for Predicting Supervisor Ratings Using the GCT, ARI and MECH as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
ARI	.306	$\hat{y} = 10.48(x_2) - 444.64$
ARI + MECH	.319	$\hat{y} = 9.64(x_2) + 3.17(x_3) - 543.29$
ARI + MECH + GCT	.324	$\hat{y} = 1.59(x_1) + 8.77(x_2) + 3.23(x_3) - 570.73$

* x_1 = GCT
x_2 = ARI
x_3 = MECH

APPENDIX B

Regression Equations for Predicting Criterion Performance
from the Miniature Aptitude Scores

KEY:

EUN= Equipment Use and Nomenclature Test
GC= Gasket Cutting Test
MR= Meter Reading Test
TR= Troubleshooting Test
EO= Equipment Operation Test
ASS= Assembly Test

Preceding page blank

Table B-1

Regression Equation for Predicting Criterion Performance on the BMF Test Using EUN, GC, TR, EO, ASS as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
ASS	.224	$\hat{y}= 3.74 + .11(x_6)$
ASS + EO	.276	$\hat{y}= .93 + .05(x_5) + .11(x_6)$
ASS + EO + GC	.295	$\hat{y}= .01 + .07(x_2) + .05(x_5) + .11(x_6)$
ASS + EO + GC + TR	.306	$\hat{y}= .08(\ _2) - .03(x_4) + .05(x_5) + .12(x_6) - .09$
ASS + EO + GC + TR + EUN	.325	$\hat{y}= 1.2 - .08(x_1) + .08(x_2) - .04(x_4) + .06(x_5) + .12(x_6)$

*x_1= EUN
x_2= GC
x_4= TR
x_5= EO
x_5= ASS

F- level insufficient for computation of MR variable.

Table B-2

Regression Equations for Predicting Criterion Performance on PV Test Using EUN, MR, TR, and EO as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
TR	.179	$\hat{y}= 7.90 + .10(x_4)$
TR + EO	.201	$\hat{y}= 5.62 + .09(x_4) + .04(x_5)$
TR + EO + EUN	.212	$\hat{y}= 4.18 + .07(x_1) + .10(x_4) + .04(x_5)$
TR + EO + EUN + MR	.219	$\hat{y}= 3.80 + .08(x_1) + .04(x_3) + .10(x_4) + .03(x_5)$

*x_1= EUN
x_2= MR
x_3= TR
x_4= EO

F- level insufficient for computation of other variables.

Table B-3

Regression Equations for Predicting Criterion Performance on the FKU,Test Using EUN, GC, TR and EO as Predictors (N= 54)

Variable	Multiple R	Regression Equation*
GC	.285	$\hat{y}$= 12.40 + .52(x_2)
GC + EUN	.384	$\hat{y}$= 3.10 + .44(x_1) + .53(x_2)
GC + EUN + TR	.420	$\hat{y}$= .53(x_1) + 50(x_2) + .16(x_4) - 1.06
GC + EUN + TR + EO	.434	$\hat{y}$= 3.05 + .56(x_1) + .50(x_2) + .20(x_4) + .09(x_5)

*x_1= EUN
x_2= GC
x_4= TR
x_5= EO

F- level insufficient for computation of other variables.

Table B-4

Regression Equations for Predicting Criterion Performance on the Sequential Test Using GC, MR, TR, EU, ASS as Predictors (N= 54)

Variables	Multiple R	Regression Equations*
ASS	.237	$\hat{y}$= 17.23 - .24(x_6)
ASS + GC	.319	$\hat{y}$= 13.19 + .29(x_2) - .25(x_6)
ASS + GC + EO	.348	$\hat{y}$= 8.49 + .29(x_2) + .08(x_5) - .25(x_6)
ASS + GC + EO + TR	.370	$\hat{y}$= 8.19 + .30(x_2) - .09(x_1) + .10(x_5) - .23(x_6)
ASS + GC + EO + TR + MR	.396*	$\hat{y}$= 5.95 + .32(x_2) + .15(x_3) - .12(x_4) + .09(x_5) - .17(x_6)

*x_2= GC
x_3= MR
x_4= TR
x_5= EO
x_6= ASS

F- level insufficient for computation or the other variable.

Table B-5

Regression Equations for Predicting Criterion Performance on the WW Test Using EUN, GC, MR, EO and ASS as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
TR	.351	$\hat{y} = .22(X_4) - .04$
TR + EUN	.402	$\hat{y} = .24(X_1) + .26(X_4) - 5.62$
TR + EUN + ASS	.430	$\hat{y} = .21(X_1) + .25(X_4) + .14(X_6) - 7.99$
TR + EUN + ASS + GC	.448	$\hat{y} = .21(X_1) + .16(X_2) + .24(X_4) + .14(X_6) - 10.66$
TR + EUN + ASS + GC + MR	.454	$\hat{y} = .21(X_1) + .16(X_2) + .07(X_3) + .22(X_4) + .14(X_6) - 11.54$
TR + EUN + ASS + GC + MR + EO	.455	$\hat{y} = .22(X_1) + .16(X_2) + .07(X_3) + .23(X_4) - .02(X_5) + .13(X_6) - 10.84$

*X_1 = EUN
X_2 = GC
X_3 = MR
X_4 = TR
X_5 = EO
X_6 = ASS

Table B-6

Regression Equations for Prediction Criterion Performance on the MR-MW Test Using EUN, GC, MR, TR, EO, and ASS as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
EO	.339	$\hat{y} = 38.47 + .29(x_5)$
EO + TR	.405	$\hat{y} = 37.60 + .24(x_4) + .24(x_5)$
EO + TR + EUN	.459	$\hat{y} = 28.90 + .44(x_1) + .33(x_4) + .21(x_5)$
EO + TR + EUN + GC	.486	$\hat{y} = 24.46 + .44(x_1) + .33(x_2) + .31(x_4) + .21(x_5)$
EO + TR + EUN + GC + MR	.497	$\hat{y} = 22.85 + .47(x_1) + .35(x_2) + .16(x_3) + .29(x_4) + .20(x_5)$
EO + TR + EUN + GC + MR + ASS	.506	$\hat{y} = 19.56 + .45(x_1) + .34(x_3) + .21(x_3) + .27(x_4) + .20(x_5) + .15(x_6)$

*x_1 = EUN GC
x_2 = GC
x_3 = MR
x_4 = TR
x_5 = EO
x_6 = ASS

Table B-7

Regression Equations for Predicting Criterion Performance on the MN-TR Test Using EUN, GC, MR, TR, EO, ASS as Predictors (N= 54)

Variables	Multiple R	Regression Equation*
EUN	.125	$\hat{y} = 11.80 + .09(x_1)$
EUN + MR	.188	$\hat{y} = 10.61 + .11(x_1) + .07(x_3)$
EUN + MR + EO	.217	$\hat{y} = 12.44 + .11(x_1) + .09(x_3) + .06(x_5)$
EUN + MR + EO + TR	.248	$\hat{y} = 11.73 + .15(x_1) + .08(x_3) + .05(x_4) + .05(x_5)$
EUN + MR + EO + TR + GC	.259	$\hat{y} = 10.95 + .15(x_1) + .06(x_2) + .08(x_3) +$.[illegible] $(x_4) + .05(x_5)$
EUN + MR + EO + TR + GC + ASS	.261	$\hat{y} = 10.45 + .14(x_1) + .06(x_2) + .09(x_3) + .05(x_4) + .05(x_5) + .02(x_6)$

*x_1= EUN
x_2= GC
x_3= MR
x_4= TR
x_5= EO
x_6= ASS

Table B-8

Regression Equations for Predicting Supervisor Ratings Using UN, GC, MR, TR, EO and ASS as Predictors (N= 54)

Variable	Multiple R	Regression Equation*
ASS	.224	$\hat{y} = 256.63 + 10.95(x_6)$
ASS + MR	.337	$\hat{y} = 12.26(x_1) + 15.49(x_6)$
ASS + MR + EUN	.389*	$\hat{y} = 12.28(x_1) + 13.81(x_3) + 15.91(x_6) - 772.06$
ASS + MR + EUN + EO	.414*	$\hat{y} = 13.17(x_1) + 15.33(x_3) - 4.09(x_5) + 15.58(x_6) - 575.49$
ASS + MR + EUN + EO + TR	.426*	$\hat{y} = 10.91(x_1) + 16.36(x_3) - 3.97(x_4) - 3.37(x_5) + 16.9(x_6) - 555.29$
ASS + MR + EUN + EO + TR + GC	.431*	$\hat{y} = 10.94(x_1) + 4.37(x_2) + 16.59(x_3) - 4.21(x_4) - 3.41(x_5) + 16.91(x_6) - 615.55$

*x_1 = EUN
x_2 = GC
x_3 = MR
x_4 = TR
x_5 = EO
x_6 = ASS